

2002

# The effectiveness of managerial leadership development programs: a meta-analysis of studies from 1982-2001

Doris B. Collins

*Louisiana State University and Agricultural and Mechanical College, dcollin@lsu.edu*

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool\\_dissertations](https://digitalcommons.lsu.edu/gradschool_dissertations)



Part of the [Human Resources Management Commons](#)

---

## Recommended Citation

Collins, Doris B., "The effectiveness of managerial leadership development programs: a meta-analysis of studies from 1982-2001" (2002). *LSU Doctoral Dissertations*. 2461.

[https://digitalcommons.lsu.edu/gradschool\\_dissertations/2461](https://digitalcommons.lsu.edu/gradschool_dissertations/2461)

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).

THE EFFECTIVENESS OF MANAGERIAL LEADERSHIP DEVELOPMENT  
PROGRAMS: A META-ANALYSIS OF STUDIES FROM 1982-2001

A Dissertation

Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

in

The School of Human Resource Education and Workforce Development

By  
Doris B. Collins  
B.S., West Virginia University, 1966  
M.S., West Virginia University, 1970  
August 2002

© 2002

Doris B. Collins

All rights reserved

## **DEDICATION**

This dissertation is dedicated to the memory of my parents, Carlie and Susie Bowers, for instilling in me the values of hard work, a good attitude and persistence, and for stressing the value of education. Their love, concern and pride in my work were always a major source of strength to me. Their encouragement, support and personal sacrifices made an everlasting impression on my life. I am humbled by my doctoral experience, because the real wisdom and ability was passed on from the two of them.

This dissertation is also dedicated to the memory of my sister, Betty Jo Maris, for being my role model, and instilling in me the inspiration to set high goals and the confidence to achieve them.

Each gave me the support, confidence, and love over the years to reach for my dreams. I know they are cheering from the balconies of heaven.

## **ACKNOWLEDGEMENTS**

This dissertation is the pinnacle of a lifelong process of learning and growing that would have been difficult without the guidance, assistance, and caring of a number of individuals. My first grade teacher, Mr. Leslie Keister, believed in my skills and abilities and set high standards of excellence for my performance. And that standard continued with my major professor and dissertation chair, Dr. Elwood F. Holton, III, who also challenged me to reach beyond what I thought I was ever capable of accomplishing.

Words cannot express my sincere gratitude and appreciation to Dr. Holton, who helped launch and sustain this journey. He challenged my thinking, writing, and research skills, and served as a mentor and coach throughout this research process. He provided unparalleled opportunities for learning, thinking, self-discovery, and growth.

My dissertation committee deserves recognition for their time, valuable suggestions, and support: Dr. Michael Burnett, for his expertise in research and his willingness to read and make constructive suggestions on this dissertation; Dr. Sharon Naquin, for coding a sample of studies, but most importantly being there with words of encouragement, providing continuous inspiring comments when I needed them most; Dr. Reid Bates, for his kind words of support of my undertaking a meta-analysis; and, Dr. Kevin Mossholder, for his insight and sound advice on the research process. Each of them, in their own way, helped make this a valuable learning experience.

I want to thank my employer, Louisiana State University, for making this educational opportunity available. The members of my doctoral cohort are also appreciated for their constant words of encouragement throughout this entire research

project: Debora Baker, Mary Boudreaux, Mary Feduccia, Janis Lowe, Lynda Wilson, and Kit Arnett.

But, no one deserves more thanks than my husband Ken whose support, encouragement, patience, and belief in me ultimately made this Ph.D. possible. He, more than any other person, has enabled me to recognize my potential. His support and companionship were critical ingredients in providing me the strength and persistence necessary to complete this work.

## TABLE OF CONTENTS

DEDICATION .....	iii
ACKNOWLEDGMENTS .....	iv
LIST OF TABLES .....	ix
ABSTRACT .....	xi
CHAPTER	
1 INTRODUCTION .....	1
Rationale for Study .....	3
Problem Statement .....	11
Purpose of Study .....	12
Objectives .....	13
Significance of the Study .....	13
2 REVIEW OF LITERATURE .....	16
Leadership Literature .....	16
Burke and Day (1986) Meta-Analysis .....	18
Related Meta-Analyses .....	22
Leadership Theories .....	25
Leadership Theories Since 1986 .....	25
Leadership Theories Prior to 1986 .....	32
Management vs. Leadership .....	36
Leadership Development Literature .....	38
Performance Improvement Literature .....	44
Outcomes of Managerial Leadership Development .....	51
Evaluation of Managerial Leadership Development .....	53
Conclusion .....	64
3 METHODOLOGY .....	66
Identification of Research Topic .....	67
Criteria for Inclusion in Study .....	68
Identification and Retrieval of Studies .....	69
Computerized Search of Databases .....	69
Manual Search of Literature .....	70
Search for Unpublished Studies .....	70
Sorting and Merging of the Pertinent Literature .....	72
Coding Research Findings and Study Characteristics .....	73
Coding Form .....	73
Coding Information .....	75
Coding Instructions .....	77
Coding Reliability .....	77
Key Coding Characteristics and Definitions .....	78

Intervention Types .....	78
Content Focus of Interventions .....	80
Managerial Leadership Development Outcomes .....	83
Study Designs .....	85
Posttest Only With Control Design (POWC) .....	87
Pretest-Posttest With Control Design (PPWC) .....	88
Single Group Pretest-Posttest Design (SGPP) .....	89
Data Reduction Techniques .....	91
Unit of Analysis .....	91
Software for Analysis .....	93
Effect Size Estimates .....	93
Interpretation of Effect Sizes .....	94
Correction for Statistical Artifacts .....	95
Sampling Error Variance .....	96
Error of Measurement .....	97
Range Restriction .....	98
Other Potential Artifacts .....	99
Outliers .....	99
Fixed Versus Random Effects .....	99
Analysis of Moderating Variables .....	100
Analysis of Heterogeneity .....	101
Summary .....	103
 4 RESULTS .....	 105
The Meta-Analysis Data Base .....	105
Profile of Managerial Leadership Development Studies .....	109
Publication Characteristics .....	109
Job Classification Level .....	110
Organization Type .....	111
Country In Which Intervention Occurred .....	112
Content Focus of Formal Training Interventions .....	112
Outcome Levels of Formal Training Interventions .....	113
Measurement Methods .....	114
Implication of Research Design in Meta-Analysis .....	115
Posttest Only With Control Group (POWC) Meta-Analysis .....	119
POWC Effect Sizes .....	121
Pretest-Posttest with Control Group (PPWC) Meta-Analysis .....	122
PPWC Effect Sizes .....	124
Single Group Pretest-Posttest (SGPP) Meta-Analysis .....	125
SGPP Effect Sizes .....	126
Moderator Analysis .....	129
POWC Study Moderators .....	131
PPWC Study Moderators .....	134
SGPP Study Moderators .....	134
Research Design as a Moderator .....	138
Objective versus Subjective Outcomes as a Moderator .....	140



Summary .....	141
5 DISCUSSION AND CONCLUSIONS .....	144
Research Questions .....	145
Research Question 1: Across Studies Measuring System Outcomes, How Effective Is Managerial Leadership Development? .....	145
Research Question #2: Across Studies Measuring Financial Outcomes, How Effective is Managerial Leadership Development? .....	150
Research Question #3: Across Studies Measuring Expertise Outcomes, How Effective Is Managerial Leadership Development? .....	152
Research Question #4: Across Studies Measuring Knowledge Outcomes, How Effective Is Managerial Leadership Development? .....	159
Research Question #5: What Moderator Effects Can Be Detected For The Following Variables: Training Content, Organization Type, Job Classification Level, Publication Type, Measurement Method, Research Design, And Objective-Subjective Outcomes?.....	161
Limitations of the Study .....	164
Implications for Practice .....	166
Implications for Future Research .....	170
Summary .....	175
REFERENCES .....	177
APPENDIX	
A SEARCH FOR UNPUBLISHED STUDIES.....	193
B META-ANALYSIS SAMPLE .....	194
C CODING FORM.....	201
D CODING INSTRUCTIONS .....	204
VITA .....	209

## **ABSTRACT**

Many organizations are committing to education and training that deepens skills, perspectives, and competencies of their leaders. This research located 103 studies from 1982-2001 with a full range of managerial leadership development interventions including feedback, developmental relationships, on-the-job experiences, and formal training. It integrated results of 83 of these studies with formal training interventions via meta-analytic techniques to determine the effectiveness of interventions, in their enhancement of performance, knowledge, and expertise at the individual, team or group, or organizational level. The studies were viewed through a “macro-lens,” that used a full range of managerial leadership development interventions (McCauley, Moxley, & Van Velsor, 1998), a high-performance leadership competency model (Holton & Naquin, 2000), and the Results Assessment System (Swanson & Holton, 1999).

The studies were separated into four separate data sets by the research design used in individual studies (posttest only with control group, pretest-posttest with control group, single group pretest-posttest, and correlation) with the unit of analysis being the outcome measure of the study. Effect sizes, derived by using Carlson and Schmidt’s (1999) formulas, were adjusted for sampling error and error of measurement. Hunter and Schmidt’s (1990) method of partitioning observed effect size variability into portions attributable to subject-level sampling error and between-study differences was used to determine the presence of seven possible moderating variables.

This research found formal training programs with knowledge outcomes highly effective. The average effect size for knowledge outcomes ranged from .96 (control group, knowledge-objective) to 1.37 (pretest-posttest, knowledge-objective). The average

effect size for expertise outcomes ranged from .30 (control group, expertise-subjective) to 1.01 (pretest-posttest, expertise-objective). System outcomes had an average effect size of .39 (control group). Two primary methodological issues were raised regarding Burke and Day's (1986) meta-analysis on the effectiveness of managerial training: 1) independence of outcomes measured (effect sizes), and 2) over weighting of studies with multiple effect sizes. Implications were provided for future research opportunities and for practical use of the findings. This meta-analysis synthesized existing studies from a broad range of settings, researchers and circumstances and integrated conflicting findings to establish a general knowledge base about managerial leadership development.

## **CHAPTER 1: INTRODUCTION**

Many organizations have become concerned about leadership inadequacies of their employees, and as a result, are committing to education and training that deepen the skills, perspectives, and competencies of their leaders (Conger & Benjamin, 1999). Gibler, Carter, and Goldsmith (2000) predicted that annual budgets for leadership development would continue to grow throughout the next decade as companies “recognize the shortage of talented managers, the importance of developing ‘bench strength’, and the need to widen perspectives in order to compete globally” (p. xii). However, Klenke (1993) pointed out that leadership education is still lagging far behind the demand curve for leaders. Leadership education is “no longer focused on the individual learner but increasingly on shaping the worldviews and behaviors of cohorts of managers and, ... transforming even entire organizations” (Conger & Benjamin, 1999, p. xii).

Even though leadership development interventions are pervasive, research indicates that organizations are spending little time evaluating the effectiveness of their interventions and, more specifically, evaluating whether those programs improve the organization’s performance (Sogunro, 1997). That leadership development efforts will result in improved leadership skills appears to be taken for granted by many corporations, professional management associations, and consultants. In essence, many companies naively assume that leadership development efforts improve organizational efforts. Leadership development is defined as “every form of growth or stage of development in the life cycle that promotes, encourages, and assists the expansion of knowledge and

expertise required to optimize one's leadership potential and performance" (Brungardt, 1996, p. 83).

There are many opinions as to why organizations are not evaluating or reporting the results of their leadership development interventions. First, the competencies required to be an accomplished leader in a high-performance organization are complex and overlapping (Collins, Lowe, & Arnett, 2000). The complexity of today's organization dictates the need for different leadership skills (Sogunro, 1997) as an endless stream of problems force leaders to allocate attention and resources among competing demands (Levi & Mainstone, 1992). Peter Vaill (1990) used the metaphor of "permanent white water" to represent the uncertainty and chaos inherent in today's managerial environment.

Second, McCauley, Moxley, and Van Velsor (1998) suggested a full range of leadership development experiences, including mentoring, job assignments, feedback systems, on-the-job experiences, and developmental relationships which include exposure to senior executives, leader-follower relationships, and formal training. While tasks and challenges encountered on the job are a major source of learning, the reality is that all jobs are not developmentally equal (McCauley & Brutus, 1998), nor can they be expressed in an objective manner, which perhaps makes evaluation more difficult.

Third, organizations appear to believe that improving knowledge and skills of the individual employee automatically enhances the organization's effectiveness. What are measured almost exclusively are the interpersonal skills and the work performance of individual managers (Moxnes & Eilertsen, 1991). Measurement of organizational

effectiveness is somewhat more difficult as it often involves analysis at multiple levels of the organization (Rummler & Brache, 1995).

Fourth, some researchers believe that evaluative studies of leadership development are sparse because of the lack of an evaluation model that adequately measures the effect of the interventions on the performance of the organization (Alliger & Janak, 1989; Bassi, Benson, & Cheney, 1996; Clement, 1982; Holton, 1996; Moller & Mallin, 1996; Newstrom, 1995; Swanson, 1998). Kirkpatrick (1998) provided a model that has been used by organizations for forty years as the tool to evaluate the effectiveness of managerial training. “The power of Kirkpatrick’s model is its simplicity...” (Alliger & Janak, 1989, p. 331). It has been used primarily to evaluate reactions, learning, and behavior, all of which are measurements of transfer of training to individual employees. However, Kirkpatrick’s model does not appear to be effective in measuring organizational performance, the effectiveness of an organization in achieving outcomes as identified by its strategic goals, or the realization of a return on investments (Holton, 1999).

### **Rationale for Study**

Leadership development interventions have gained more attention because most organizations are facing a multitude of outcome-based demands on their time and resources – demands that stem from a variety of driving forces including federal mandates, increased global competition, and national accreditation standards. Leadership development is also important in global organizations because of dual reporting structures, proliferation of communication channels, overlapping responsibilities, and barriers of distance, language, time, and culture (Friedman, 2000). Given that it is the

people within organizations who sell and market, develop and create products, make decisions and implement programs, the development of human resources is especially vital to the success of the global organization (Caligiuri & Stroh, 1995). But, very few leadership development interventions have a global focus (Marquardt & Engel, 1993).

Because of the demands on today's organizations, and the need to justify training programs to top management, human resource practitioners have realized that organizational outcomes must be the driving force and energy behind the design and implementation of all leadership development interventions. According to Swanson and Holton (1999), "every (leadership development) intervention should lead to a system outcome at some point" (p. 69).

Most organizations recognize that effective leadership is one of the most powerful competitive advantages they can possess. HRD scholars make the case "that people are the only assets with the creativity and adaptive power to sustain an organization's success in today's dynamic business world" (Krohn, 2000, p. 63). Torraco and Swanson (1995) further stated that investment in "employee education and training increasingly funds the development of an infrastructure to support the sustainable competitive advantage that a highly-trained workforce provides" (p. 13). Therefore, "it is the development of workplace expertise that is becoming vital for organizations to adapt to change and maintain optimal organizational performance" (Herling, 2000, p. 9).

A review of the leadership development literature illustrated that the relationship between leadership development and organizational performance is not clear. Burke and Day (1986) conducted a meta-analysis of 70 published and unpublished studies spanning over 30 years to study the effectiveness of managerial training and found that only two of

the studies used organizational variables as outcome criteria. They found that the primary purpose of most leadership development interventions was to improve individual managerial skills and on-the-job performance. They also found that some cases had mixed results in demonstrating that leadership development interventions enhance individual, group, and/or organizational effectiveness. (See Chapter 2 for the details of Burke and Day's study.)

No comprehensive study of managerial leadership development interventions has been published since Burke and Day's (1986) cumulative study on managerial effectiveness. Since Burke and Day's study, the focus of leadership has shifted to strategic leadership because of increasing global competitiveness; the almost continuous restructuring activities; demographic changes in the workforce; customer demands; and rapid technological changes (Gibler, Carter, & Goldsmith, 2000; Hooijberg, Hunt, & Dodge, 1997). More importantly, the ability of multinational companies to compete in the global market is now contingent upon their ability to change and adapt resources strategically to the nuances of the broadening global playing field (Caligiuri & Stroh, 1995). Globalization requires that leaders delegate decision-making, geographically disperse key functions across units of different countries, de-layer at the organizational level, de-bureaucratize formal procedures, and differentiate work responsibilities and authority across networked subsidiaries (Friedman, 2000).

In addition, since 1986 new leadership practices and theories such as transformational leadership (Bass, 1985), team leadership (Hackman & Walton, 1986; Larson & LaFasto, 1989), 360-degree feedback (Lepsinger & Lucia, 1997), and on-the-job experiences (McCauley & Brutus, 1998) have been introduced into the literature and



have spurred new content areas in leadership development research. “At the same time, the world of executive and management education has been undergoing an equally radical shift in learning approaches and program design” (Conger & Benjamin, 1999, p. xiv). Learning methods are more dynamic, and they allow learners to address and learn from real challenges that help them resolve real issues. And, “there has been greater emphasis placed on groups of managers” (p. 20).

Therefore, the focus of leadership development research today is considerably different than what Burke and Day (1986) found because of these new leadership practices and theories, and a shift in management and leadership practices. Some organizations have constructed a framework that integrates leadership development interventions, strategic goals and objectives, and other human resource systems (McCall, 1998). In these organizations the implication is that leadership development interventions “must be anchored in and driven by business strategy” (McCall, 1998, p. 17). According to Friedman (2000), “management always matters, but in this more complex and fast-paced system, management and strategic vision matter a lot more” (p. 231). Therefore, organizations have discovered that it is important to align the goals for management or leadership development interventions with the strategic vision of the organization, and to train managers in their new roles in strategic management.

In addition to the shift in leadership theories and practices since Burke and Day’s study (1986), there is also an on-going controversy in the leadership literature regarding the distinction between leadership and managerial behaviors. Many individuals often have difficulty drawing a line between the two. Since 1977, conventional wisdom has been that leaders and managers are different types of people, with managers most

concerned with how things get done and leaders with building commitment and vision (Kotter, 1990). A more recent way of characterizing the distinction between management and leadership is to contrast transformational and transactional leadership (Bass, 1985). As Yukl (1989) said, “Nobody has proposed that managing and leading are equivalent, but the degree of overlap is a point of sharp disagreement (p. 253).”

This research adopted the term “managerial leadership development” to integrate the functions described in the literature depicting leaders and managers (Bass, 1990; Fleischman et al., 1991; House & Aditya, 1997; Kotter, 1990; Yukl, 1989; Yukl & van Fleet, 1992). It also adopted the full range of leadership model (Avolio, 1999; Bass, 1998) where leaders and manager’s behaviors are different, but all leaders displayed both types of behavior to varying degrees, and transformational leadership augments transactional leadership. Therefore, the term “managerial leadership” integrated the traditional managerial and leadership behaviors where those behaviors are different but complementary. Source (1994) stated, “managerial leadership is indeed a subtle mixture of formal authority, skills, knowledge, information, intelligence, courage, tenacity, instinct and hard work” (p. 3). As individuals rise to higher levels of formal leadership in organizations, the balance between leader and manager behavior shifts, but there are very few instances where a person can develop leadership skills without also being competent at managerial functions (Holton & Naquin, 2000).

This research subscribed to Holton and Naquin’s (2000) definition of high-performance leadership as “leading and managing people and organizational systems to achieve and sustain high levels of effectiveness by optimizing goals, design and management at the individual, process and organizational levels” (p. 1). Effective

managerial leadership is complex, and it may involve many very different and distinct activities and actions by the leader-manager on the specific intervention, attitude, or trait that causes the manager to have influence on subordinates (Sourcie, 1994; Yukl, 1989).

Integrating the results of leadership and management development research via meta-analytic techniques (Hunter & Schmidt, 1990; Lipsey & Wilson, 2001) will assist in determining the effectiveness of managerial leadership development interventions, in their enhancement of organizational performance, individual knowledge, and expertise. Little is known about what knowledge and skills or processes in managerial leadership development interventions contribute to organizational performance (Campbell, Dunnette, Lawler, & Weick, 1970; Fiedler, 1996; Lynham, 2000). Therefore, this research focused on outcomes, or the effectiveness of managerial leadership development programs, in terms of knowledge, expertise, financial or system results at the individual, team or group, or organizational level (Rummler & Brache, 1995). An *outcome* was defined for this research as “a measurement of effectiveness or efficiency (of the organization) relative to core outputs of the system, subsystem, process, or individual” (Holton, 1999, p. 33).

Managerial leadership development outcomes have traditionally focused on individual learning and skills without regard to organizational performance. Those who have evaluated outcomes have traditionally used Kirkpatrick’s evaluation model (1998). However, according to Holton (1996), “a shortcoming of the Kirkpatrick model is that it does not fully identify all constructs underlying the phenomena of interest, thus making evaluation impossible (p. 6). Clement (1982) also noted that essential elements are not present in Kirkpatrick’s model and it is questionable whether the model can be applied

universally. “The missing elements and relationships prohibit making accurate statements about system states, developing propositions, and hypotheses, and making predictions” (Holton, 1996, p. 7). For these reasons, the Results Assessment System (Swanson & Holton, 1999) was used in this research to analyze the outcomes of leadership development studies from both a learning and performance perspective.

The Results Assessment System (Swanson & Holton, 1999) enables practitioners to measure results within three domains: performance, learning, and perception, each of which has two options. Performance results are either system or financial results. System results are defined as “the units of mission-related outputs in the form of goods and/or services having value to the customer and that are related to the core organizational, work processes, and group or individual contributors in the organization” (p. 14). Financial results are defined as “the conversion of the output units of goods and/or services attributable to the intervention into money and financial interpretation” (p. 16).

Performance-level assessment requires that mission-related performance outcomes be connected to the mission of the system (Swanson & Holton, 1999). Performance outcomes typically are assessed in terms of being counted or time taken to produce the units of goods or services. Swanson and Holton clarified that performance outcomes are classified in terms of “the performance levels at the whole system level (organization), the work processes within the system (subsystem), or the contributor level (individuals or intact work groups)” (p. 64). According to Swanson and Holton, “within the performance domain, ...a complex organization can have a variety of performance outcomes” (p. 64) but a “unit of performance must be selected as the focal point of the assessment” (p. 67).

Learning results as defined by the Results Assessment System (Swanson & Holton, 1999) are delineated into expertise or knowledge results. Expertise results are defined as “human behaviors having effective results and optimal efficiency, acquired through study and experience within a specialized domain” (p. 17). Human expertise is the most complex of learning results. The premise is that people with expertise have knowledge and are able to act on that knowledge (Swanson & Holton, 1999). Measuring human expertise requires that an individual demonstrate his or her behavior in a real or simulated setting.

Knowledge results are defined as “mental achievement acquired through study and experience” (Swanson & Holton, 1999, p. 17). Swanson and Holton believed that knowledge, an intellectual or cognitive result of learning experiences, was the basic learning result of an intervention. Measures of knowledge confirmed the level of knowledge held by individuals within a particular subject area.

Perception results are a third category of outcomes defined by the Results Assessment System (Swanson & Holton, 1999). Perception outcomes are described as participant perceptions and stakeholder perceptions. The perception results category has the least valid information about outcomes, and therefore was not used in the macro-lens for this research.

The high-performance leadership competency model (Holton & Naquin, 2000) was used in the research to define intervention content areas. The high-performance leadership competency model integrated disparate streams of leadership research into a broad holistic model for developing high-performance leadership competencies. The model was developed by an examination of leadership through an organizational

performance lens. The organizational performance lens was important to this research as it focused on all levels of the organization, the organizational, process, and individual performance levels as defined by Rummler and Brache (1995). In addition, the high-performance leadership competency model was grounded in research and theory and translated research and theory on leadership into a practical framework suitable for managerial leadership development curriculum design.

This meta-analysis viewed leadership and management development studies from 1982-2001 through a “macro-lens,” a slightly different approach than used in Burke and Day’s meta-analysis (1986). This lens used the full range of managerial leadership development interventions (McCauley, Moxley, & Van Velsor, 1998) to define the intervention type and the high-performance leadership competency model (Holton & Naquin, 2000) to define the content area of the intervention. The Results Assessment System (Swanson & Holton, 1999) in conjunction with the objective-subjective outcomes described by Burke and Day was used to define the outcome category.

### **Problem Statement**

The nature of management and leadership has changed significantly and organizations are experiencing an increased number of outcome-based demands on their time and resources. Organizations also are committing to an increased number of managerial leadership development interventions and take for granted that those interventions enhance their organization’s effectiveness. But, there remains a void as to what is known about managerial leadership development and the contribution of managerial leadership development interventions to individual knowledge and expertise as well as organizational performance (Lynham, 2000).

No meta-analysis has been published on the effectiveness of managerial leadership development studies since Burke and Day's (1986) meta-analysis. Burke and Day's research produced mixed results on the effectiveness of managerial training. In addition, only two out of 70 studies had organizational performance as the outcome variable. A preliminary study of leadership development interventions from 1986-2000 conducted by Collins (2002) located 18 out of 54 studies with organizational performance as the outcome variable and four of these studies had negative results. With an apparent increase in organizational level studies and organizations taking for granted that leadership development programs enhance organizational performance, the researcher wondered if managerial leadership development interventions truly are effective.

A cumulative study of the outcomes of managerial leadership development interventions is needed to identify the intervention content areas or types of managerial leadership development interventions that enhance individual, group or team, or organizational effectiveness. The findings of this research should be of theoretical interest to researchers as well as of practical use to organizational decision-makers. Organizations will hopefully use results of this type of research to design more effective managerial leadership development interventions and justify the return on their training investment.

### **Purpose of Study**

This study applied the process of meta-analysis to the available managerial leadership development research from 1982-2001, after Burke and Day's (1986) meta-analysis, to determine the effectiveness of managerial leadership development interventions. The study determined the effectiveness of managerial leadership development interventions by researching the association between system, financial,

expertise, and knowledge results (Swanson & Holton, 1999) in relation to the intervention content area and the type of intervention.

### **Objectives**

For each of the four research design types (posttest only with control group, pretest-posttest with control group, single group pretest-posttest, and correlation), the objective of this study was to answer the following research questions:

- 1) Across studies measuring system outcomes, how effective is managerial leadership development?
- 2) Across studies measuring financial outcomes, how effective is managerial leadership development?
- 3) Across studies measuring expertise outcomes, how effective is managerial leadership development?
- 4) Across studies measuring knowledge outcomes, how effective is managerial leadership development?
- 5) What moderator effects can be detected for the following variables: training content, organization type, job classification level, publication type, measurement method, research design, and objective-subjective outcomes?

### **Significance of the Study**

Managerial leadership development is a young field for which little information is available in the literature, particularly relative to outcomes of the organization as a system. Researchers have continuously expressed that more empirical studies are needed for a thorough examination of managerial leadership development (Burke & Day, 1986; Brungardt, 1996; Lynham, 2000). This meta-analysis of the literature provided a



synthesis that can potentially make significant contributions to both the general knowledge about managerial leadership development and to HRD practice.

This meta-analysis provided an overview of the research on managerial leadership development from 1982-2001, as it synthesized the existing studies over a broad range of settings, researchers, and circumstances to provide a new insight into the topic. This research focused on both learning and performance -- knowledge, expertise, financial and system results at the individual, group or team, or organizational level -- as outcomes of managerial leadership development interventions. However, the need in the leadership development research arena was not only for more empirical data, but also for some means of making sense of the data that had been accumulated. Therefore, this meta-analysis on the effectiveness of managerial leadership development interventions is important to theory development and practical problem solving because it integrated conflicting findings to establish a general knowledge base about managerial leadership development. Meta-analysis provided a powerful statistical technique to eliminate sampling error, correct for unreliability, and provide for generalizability of studies to the effectiveness of all managerial leadership development interventions.

There is more pressure today than ever before to produce results with HRD programs. Whether the result of globalization, competition, the economy, or merely a lack of understanding of the importance of human resources, top management in many organizations is questioning the return on investment for developmental interventions (Dionne, 1996; Holton, 1996). As a result, HR practitioners often have to justify the financial contribution of their managerial leadership development interventions and explain how the interventions add value to the company (Gordon, 1987; Plant & Ryan,

1992). Unfortunately, many companies with extensive training budgets do not have a comprehensive approach to tying their training investment to bottom-line results (Phillips, 1997). Researchers believe that significant financial payoffs are found among companies that emphasize training and development (Huselid, 1995; Jacobs & Jones, 1995; Lam & White, 1998; Swanson, 1994; Ulrich, 1997), and that organizations with stronger learning environments demonstrate greater organizational effectiveness (Tannenbaum, 1997). This meta-analysis integrated managerial leadership development studies with conflicting theories to provide practitioners with information regarding the effectiveness of managerial leadership development interventions.

## **CHAPTER 2: REVIEW OF LITERATURE**

This chapter begins with a review of the literature on leadership and then describes Burke and Day's meta-analysis (1986) on the effectiveness of managerial training from 1951-1982, as that study serves as a prototype and a starting point for the collection of studies for inclusion in this research. Descriptions of several other meta-analyses on related topics follow the description of the Burke and Day study. The chapter continues with a review of leadership theories and a contrast of the literature on management and leadership. It continues with a review of the literature on leadership development, performance improvement theory, outcomes, and evaluation of leadership development interventions.

### **Leadership Literature**

Volumes of literature exist on the concept of leadership, and leadership researchers believe that leadership really makes a difference (Bass, 1990; Burke & Day, 1986; Clark, Clark, & Campbell, 1992; Ulrich, Zenger, & Smallwood, 1999). Leadership is defined in many ways. Lohmann (1992) defined leadership as "the formulation of a vision, developing a climate of trust within the organization, and empowering others" (p. 59). Leadership is also defined as a process whereby individuals influence groups of individuals to achieve a shared goal or commonly desired outcomes (Northouse, 1997). Rost (1991) analyzed a total of 587 works that referred to leadership in their titles and found that 366 of them did not specify a definition of leadership. From the analysis, Rost found that most leadership literature focused on leader ability, traits or behaviors.

To be in a strategic position and to be competitive, winning organizations have leaders at every level (Lercel & Field, 1998). In today's global marketplace pleasing

customers and making profits are functions of quick thinking and agility. According to Lercel and Field, the only differentiator was the quality and speed of action, and the only way to improve quality and speed was to have leaders in the organization prepared to make smart decisions and implement them efficiently at all levels.

Klenke (1993) described leadership as a field that is “riddled with paradoxes, inconsistencies, and contradictions. Ulrich et al. (1999) described the array of leadership attributes as comprised of “confusing and often overlapping terms” (p. 4). There are probably few areas...which have produced more divergent, inconsistent, overlapping definitions, theories, and educational models than leadership” (Klenke, 1993, p. 112).

Yukl (1989) described the status of the field of leadership as being “in a state of ferment and confusion. Most of the theories are beset with conceptual weaknesses and lack strong empirical support. Several thousand empirical studies have been conducted on leadership effectiveness, but most of the results are contradictory and inconclusive... The confused state of the field can be attributed in large part to the disparity of approaches, the narrow focus of most researches, and the absence of broad theories that integrate findings from the different approaches” (p. 253).

Fleishman et al. (1991) developed a taxonomy of descriptive leader behaviors from 65 authors, which provided a systematic definition of leadership behavior for use in designing leadership development interventions. The major approaches to leadership study were identified as the power-influence approach, managerial behavior approach, trait approach, situational approach (nine different ones including path-goal, situational leadership, contingency theory, and leader-member exchange) and transformational or

charismatic leadership (Yukl, 1989). These approaches are defined under the Leadership Theories section later in this chapter.

Brungardt (1996) indicated that current leadership literature focuses on three areas: general approaches to leadership, leadership development research, and, leadership education. Lynham (2000) believed that leadership development research covered four separate areas: early childhood and adolescents; role of formal education; adult and on-the-job experiences; and specialized leadership education programs. According to Lynham, the term “leadership education” was often used in the literature instead of leadership development. Leadership education is a one-time intervention while leadership development experiences contribute to an individual’s professional development over the course of their lifetime (Lynham, 2000).

Avolio (1999) described leadership as a systematic relationship where “no leader leads without followers” (p. 3). To measure leadership effectiveness, Avolio believed that the focus should be less on what the leader does and more on what the followers do. Avolio committed to a “full range” of leadership, defined by three broad categories of transformational, transactional, and laissez-faire or non-transactional leadership.

### **Burke and Day (1986) Meta-Analysis**

Burke and Day (1986) applied meta-analysis to available managerial training and development studies to determine the types of management training that were effective, to what degree they were effective, and the relative effectiveness of the different training methods in improving learning or the acquisition of skills. The meta-analysis conducted by Burke and Day is commonly regarded as the principal empirical support for the effectiveness of managerial training and leadership development programs. Yet, Lynham

(2000) believed that the field of leadership development “may be well served by further purposeful and scholarly inquiry and study” (p. 1). In addition, Lynham stated that there is a “need to gather up studies and understanding of leadership development, and to conduct analyses of the evolution and nature of what is really known in this field” (p. 5).

Burke and Day’s meta-analysis (1986) included 70 published and unpublished studies spanning from 1951-1982. Studies included in their meta-analysis involved managerial or supervisory personnel, evaluated the effectiveness of more than one training program, and included at least one control or comparison group. Burke and Day captured a variety of information regarding each case, including the training content area, training method, outcome variable, managerial level, years of work experience and sex of participants, type and length of training program, time between training and the evaluation process, and the assignment of subjects.

Burke and Day (1986) incorporated the following training content areas in their analysis: general management, human relations, self-awareness, problem solving/decision making, rater training, and motivation/values. Descriptions of those content areas were as follows:

- 1) General management training taught facts, concepts, and skills and included training topics such as labor relations, a broad focus on management theory and practice, company policies and procedures, labor economics, and general management functions.
- 2) Training in the human relations content area focused on the human relations problems of leadership, supervision, attitude toward employees, and communications.

- 3) Studies that were coded into the self-awareness training content area involved the understanding of one's behavior, how others view one's behavior, and learning one's strengths and weaknesses. Examples in the self-awareness content area were sensitivity training and transactional analysis.
- 4) Problem solving training included studies with a wide range of work problems that managers encounter including generalized problem solving and decision-making skills.
- 5) Rater training programs taught participants to minimize errors in observing and evaluating subordinates.
- 6) Motivation/values training included programs designed to increase the manager's motivation and modify manager's values or attitudes.

Burke and Day (1986) also categorized studies by training method. Training methods used were: lecture, lecture/group discussion, leader match, sensitivity training, behavioral modeling, lecture/group discussion with role-playing or practice, and multiple techniques.

In addition, four criterion-measure categories were developed by Burke and Day (1986) on the basis of two dimensions: (a) level of criterion and (b) subjectivity-objectivity. The level of criterion dimensions was determined according to Kirkpatrick's evaluation model (1976). Three of Kirkpatrick's four levels were used: learning, behavior, and results. Categories of subjective learning, objective learning, subjective behavior, and objective results were used as criterion-measure categories. Descriptions of these criterion-measure categories are as follows:

- 1) Subjective learning included measures that assessed what principles, facts, attitudes, and skills were learned during or by the end of training as communicated in statements of opinion, belief, or judgment completed by the trainee or trainer.
- 2) Objective learning was measures that assessed what principles, facts, attitudes, and skills were learned during or by the end of training by objective means, such as number of errors made or number of solutions reached, or by standardized test.
- 3) Subjective behavior included measures that evaluated changes in on-the-job behavior perceived by trainees, peers, or supervisor.
- 4) Objective results were measures that evaluated tangible results, such as reduced costs, improved quality or quantity, promotions, and reduced number of errors in making performance ratings.

Burke and Day's (1986) conclusion, based on 70 studies, was "that managerial training is, on the average, moderately effective" (p. 232). Although they do not report an overall effect size across all variables, they provided true mean effect sizes (in parentheses) for each of the four criteria used: subjective learning (.34), objective learning (.38), subjective behavior (.49), and objective results (.67).

Burke and Day's (1986) study clarified the breadth of managerial training, but indicated that more empirical research was needed before conclusive statements could be made. They found that managerial training was pervasive and primarily focused on improving individual managerial skills and on-the-job performance. But, the lack of



evaluative research caused Burke and Day to believe that organizations were unaware of the effectiveness of management training programs in improving job performance.

Other significant conclusions from the Burke and Day (1986) study were:

- 1) Researchers needed to improve reports that evaluate organizational interventions to provide cumulative analyses of the effectiveness of managerial training.
- 2) Trainers and organizational decision-makers should not rely on the training program content area descriptions when choosing the utility of the managerial training program.
- 3) The level of experience of the trainer may be significant in influencing the effectiveness of the training program.
- 4) Different management training methods do not necessarily lead to increased knowledge and improved performance.

Burke and Day found that short time frames and reliance on self-report measures typified management development research. Only two of the studies (3%) in Burke and Day's meta-analysis used organizational variables as outcome criteria. Some studies in their research had mixed results in demonstrating that managerial leadership development programs enhanced individual, group, and/or organizational effectiveness. The behavioral modeling approach (Sorcher & Goldstein, 1972) and the leader-match-training program (Fiedler & Chemers, 1984) were supported by the Burke and Day study.

### **Related Meta-Analyses**

Several other meta-analyses on topics related to leadership development (Bayley, 1988; Chen, 1994; Lai, 1996; Leddick, 1987; Zhang, 1999) were located. Bayley (1988)

synthesized 79 studies between 1966 and 1985 using meta-analysis and reported highly significant effects of continuing education on behavioral change in clinical practices. Both pre- and post-test and experimental versus comparison group studies were included in Bayley's research. Bayley recommended that researchers provide more detailed and relevant information about interventions and adhere more consistently to acceptable standards of scientific reporting.

Chen (1994) used meta-analytic procedures to describe and statistically integrate 25 studies from the empirical literature regarding the effectiveness of cross-cultural training for managers. A great majority of the studies (88%) used control group design. Chen's meta-analysis produced a highly significant average effect size (1.60) for the comparisons between those who received cross-cultural training and those who did not, indicating that the average trainee was 1.60 standard deviation higher than controls on the cross-cultural training effectiveness measures. Chen discovered that control group studies produced lower effect sizes than single group pretest-posttest studies. Chen also found that the longer the time between cross-cultural training and the measurement of training effectiveness, the less effective the training was judged to be by the primary study participants, with almost 56% of the variability in effect size magnitude caused by the time of outcome measurement. The results of Chen's meta-analysis did not conclude that any certain type of cross-cultural training program was more effective than another one.

Lai (1996) conducted a meta-analysis to integrate findings of twelve studies on the program effectiveness of educational leadership training. The study features of Lai's research, which used only experimental or quasi-experimental design, consisted of four clusters: study design, participant's characteristics, training characteristics, and outcome

measurements. The findings showed that two study features, training time frame and posttest timing, had medium correlations with leadership training effects, with the two variables explaining on 14.5% of the total variance in outcome measures of leadership training. The overall effect size was .269 with a standard error of .074, indicating that educational leadership training had a small effect when leader behavior changes were measured.

Leddick (1987) analyzed 48 field experiments between 1971-1981 to explore differences in the effectiveness of training programs. Across all studies and multiple research design types, the analysis produced an overall effect size of .67, with a .98 effect size for managers only. Leddick found that effect sizes were smaller when true controls or non-equivalent control groups were used than single groups (i.e., control group effect sizes were .39 while single group effect sizes was .96 on the same dependent variable). An additional conclusion from Leddick's study was that knowledge objectives seemed to be associated with stronger productivity improvements than other types of objectives.

Zhang (1999) applied meta-analysis procedures to experimental evaluation studies to find out the magnitude of the effect of management training from 1983-1997 on trainee's learning, job performance, and organization results. The study followed Burke and Day's coding criteria and included forty-seven empirical studies on training for managerial personnel in business and industry and in education. Zhang included experimental and quasi-experimental studies, most of which were found in journal articles, while one-third were doctoral dissertations. The results indicated that evaluation was being conducted beyond the reaction and learning levels. Zhang's research produced a .47 effect size for studies with knowledge-subjective outcomes, .80 for knowledge-

objective, .50 for expertise-subjective and .49 for studies with system-objective outcomes. A major finding was that management training made a significant difference in trainees' learning when self-efficacy and various knowledge tests measured the outcomes. A human relations leadership program made a significant difference in trainee's job performance when performance appraisal instruments measured their on-the-job behavior. Management training programs were effective when measured by subjective result criterion, such as employees' commitment to the organization and job satisfaction. There was a significant difference in the training effect measured by objective organization result criterion, such as job accuracy, turnover and productivity. Zhang recommended that more quantitative reviews be conducted using meta-analysis to accumulate quantitative data of training effectiveness across studies and that more high quality empirical studies be conducted. Zhang also concluded that measurement of organization results outcomes needed more research in which the organizational indicators that are most relevant to training are prioritized.

### **Leadership Theories**

This section begins with a description of the leadership theories that have appeared in the literature since Burke and Day's (1986) meta-analysis. It continues with leadership theories prior to 1986 and a contrast of literature on management versus leadership.

#### **Leadership Theories Since 1986**

Leadership theories since Burke and Day's (1986) study are transformational leadership (Bass, 1985), team leadership (Hackman & Walton, 1986), 360-degree feedback (Lepsinger & Lucia, 1997), and on-the-job experiences (McCauley & Brutus,

1998). These leadership theories have spurred new content areas for managerial leadership development programs.

Transformational leadership. The term “transformational leadership” was coined by Downton (1973) and spurred by Burns’ (1978) concept of transforming leadership. Transformational leaders “engage with others in such a way that leaders and followers raise one another to higher levels of motivation and morality” (Burns, p. 20).

Bass (1985) developed transformational leadership theories as an acknowledgement that organizations must make major changes in the way things are done in order to survive in the face of increasing economic competition. Transformational leadership (Bass, 1985) is more clearly defined as changing organizations so they are poised to become more competitive in a global society (Porter 1985).

Transformational leaders are strategic leaders, but “little attention has been given in the literature to strategic leadership” (House & Aditya, 1997, p. 445). As the dominant leadership theory of the 1990’s, transformational leadership contends that visionary leaders are catalysts for organizational transformation (Collins et al., 2000; Klenke, 1993). Tichy and Devanna (1990) believed that the power of transformational leadership is the visualization of the organization in the future, and the ability to articulate, develop, elaborate, and share that vision. Rolls (1995) suggested that transformational leaders build awareness and acceptance of goals and mission, motivate support among organizational members for organizational goals, and are able to influence others because they create organizational meaning.

Bass (1998) claimed that transformational leadership can "move followers to exceed performance" (p. 2). Bass identified four components of transformational leadership:

- 1) Leadership is charismatic such that the follower seeks to identify with the leaders and emulate them. Transformational leaders are role models, who are admired, respected, and trusted. Followers identify with them and believe that they have extraordinary capabilities, persistence, and determination.
- 2) The leadership inspires the follower with challenge and persuasion providing a meaning and understanding. "Transformational leaders arouse team spirit, enthusiasm and optimism. "Leaders get followers involved in envisioning attractive future states; they create clearly communicated expectations that followers want to meet and also demonstrate commitment to goals and the shared vision" (p. 5).
- 3) The leadership is intellectually stimulating, expanding the follower's use of their abilities. Transformational leaders "stimulate followers' efforts to be innovative and creative by questioning assumptions, reframing problems, and approaching old situations in new ways" (p. 5). They solicit new ideas and creative problem solutions from followers, and include followers in the process of addressing problems and finding solutions.
- 4) The leadership is individually considerate, providing the follower with support, mentoring, and coaching. "Individualized consideration is practiced when new learning opportunities are created along with a supportive climate"

(p. 6). Individual differences are accepted and a two-way exchange in communication is encouraged.

Yukl (1989) described transformational leadership as being the “process of influencing major changes in the attitudes and assumptions of organizational members and building commitment for the organization’s mission, objectives, and strategies” (p. 269). The theory is further described by Yukl as a “leadership process that is recognized primarily by outcomes such as major changes in the culture and strategies of an organization or social system” (p. 269). It also involves influence by a leader on subordinates, but the effect of the influence is to empower subordinates to participate in the process of transforming the organization. Yukl further described transformational leadership as a shared process, involving actions of leaders at various levels and sub-units of an organization, not just the chief executive. The theory of charismatic and transformational leadership simultaneously involves leader traits, power, behavior, and situational variables (Yukl, 1989). Yukl defined charismatic leadership more narrowly than transformational leadership. Charismatic leadership refers to perception that a leader possesses a divinely inspired gift where followers not only trust and respect the leader, but also idolize or worship the leader as a superhuman hero.

Waldman, Ramirez, House, and Puranam (2001) realized that strategic management theory had become increasingly concerned with top-level leadership and their effects on strategy formulation and firm performance. Waldman et al. (2001) systematically examined the effects of CEO transactional and charismatic leadership on firm profitability both in environments perceived to be certain and in those perceived to be uncertain. They found that top managers and firm outcomes might depend to a large

extent on the managers' charismatic leadership, but only under conditions of perceived environmental uncertainty. In addition, transactional leadership did not add unique variance to the prediction performance beyond that of charisma (Waldman et al., 2001).

Team leadership. Hackman and Walton (1986) and Larson and LaFasto (1989) developed team leadership theories. Northouse (1997) described team leadership theory as a “focus on organizational teams, the vital functions of team leadership, and factors contributing to organizational team effectiveness” (p. xiv). He further stated that team leadership theory must “focus on what makes teams effective and what constitutes team effectiveness” (p. 163). Northouse defined teams as “organizational groups composed of members who are interdependent, who share common goals, and who must coordinate activities to accomplish these goals” (p. 159).

Leadership in organizations, groups, or work teams has become one of the most popular and rapidly growing areas of leadership theory (Daugherty & Williams, 1997; Penwell, 1992; Quast & Hazucha, 1992; Riechmann, 1992). Organizational restructuring has shifted decision-making powers downward from the traditional hierarchy to more self-managed teams, empowering them in new ways. In some situations, the leadership role even rotates among team members. Project management teams, task forces, work units, and standing committees are all examples of teams.

Avolio (1999) indicated that the prerequisite for being a team is in “sharing some common purpose” (p. 111). The full range model of leadership as described by Avolio can also be used for examining both individual and team leadership as the model represents a “cross-level” theory. Moving to a team-based or networked organization allows leaders to move back and forth between an individual and team leadership system.



The sharing of leadership, indicative of teams, becomes “feasible when commitment has been built into the team, along with members’ identification with each other, the mission, and the vision” (p. 119).

Teams often lead global organizations. Maznevski and DiStefano (2000) defined global teams as teams of managers from different parts of a multinational organization working together to achieve a team-specific mandate that is global in scope. Global teams “usually consist of members who have very different cultural and functional backgrounds and who differ in their assumptions about how to approach relationships and how to make decisions” (Maznevski & DiStefano, p. 196). Global teams provide an ideal forum in which the organization’s future global leaders can develop the knowledge and skills they need (Maznevski & DiStefano, 2000).

With the increase in organizational work teams and their expanding role in the complex and rapidly changing organizational structures, it is essential to understand the role of leadership within these teams to ensure that they teach success and to avoid team failure. “The practical necessity of understanding the nature of organizational teams and the leadership within them is forcing theory and research into new directions that offer great promise for understanding team leadership” (Northouse, 1997, p. 160). “Leaders must recognize their area of strength, weakness, and interest, and build a team that collectively covers all the bases” (Ulrich et al., 1999, p. 34).

Team leadership theory will continue to be a focus of managerial leadership development research in the future as more organizations employ management teams to accomplish core organizational goals. It is important to note that Baker, Walsh, and

Marjerison (2000) incorporated team management theory into the competencies necessary for leadership in high-performance organizations.

360-degree feedback. Many corporations have embraced multi-rater feedback (360-degree feedback) as a means of achieving business goals and enhancing employee's performance (Lepsinger & Lucia, 1997). The feedback process involves collecting perceptions about a person's behavior and the impact of that behavior from the person's boss, direct reports, colleagues, fellow project team members, and external stakeholders. Managerial leadership development studies since 1982 (Facteau, Facteau, Schoel, Russell, & Poteet, 1998; Fiedler, 1992; Hazucha, Hezlett, & Schneider, 1993; Marson & Bruff, 1992; Spreitzer, McCall, & Mahoney, 1997; Tucker, Bass, & Daniel, 1992;) indicate that 360-degree feedback has become a method of improving employee performance. Kuchinke (2000) described feedback as the general communication process in which a sender conveys a message to a recipient related to some aspect of the recipient's behavior. Feedback is critical at the individual, group, and organizational levels because it provides information about possible discrepancies between current and target performance levels and outcomes.

On-the-job experiences. Managerial leadership development through on-the-job-experiences has emerged as a powerful source of learning. McCall (1998) believed that on-the-job experiences were the primary classrooms for the development of leadership skills. These developmental jobs provide transitions that put the manager into new situations with unfamiliar responsibilities and tasks where they create change and build relationships (Brutus, Ruderman, Ohlott, & McCauley, 2000; McCall, Lombardo, & Morrison, 1988; McCauley & Brutus, 1998). According to McCall, et al. (1988), most

development of successful business leaders took place on the job and not in seminars, classrooms, and MBA programs. Additional literature regarding on-the-job experiences will be cited in the Leadership Development section below.

### **Leadership Theories Prior to 1986**

Theories developed prior to 1986 also continue to impact managerial leadership development. Northouse (1997) defined these theories as trait, style, power-influence, contingency, situational, path-goal, and leader-member exchange theories.

Trait approach. The importance of leader traits continues to resurface (Kirkpatrick & Locke, 1991) in the leadership literature. The trait approach, the primary leadership theory from 1930 – 1950, focuses on the personality of the leader as the primary leadership characteristic. House and Aditya (1997) described the trait-approach theory as “individual characteristics that differentiate leaders from non leaders” (p. 410). Northouse (1997) described the personality traits that effect leadership as intelligence, self-confidence, determination, integrity, and sociability. Yukl (1989) described the trait approach as emphasizing the personal attributes of leaders. “Early leadership theories attributed success to possession of extraordinary abilities such as tireless energy, penetrating intuition, uncanny foresight, and irresistible persuasive powers” (Yukl, 1989, p. 260). Yukl believed that the focus of most trait research had been on “managerial motivation and specific skills, whereas earlier research focused more on personality traits and general intelligence” (p. 260).

Style approach. The style (or behavioral) approach was developed through studies at Ohio State University and University of Michigan in the late 1940’s and focused on the behavior of leaders. What leaders do on the job and how they act in

various contexts are elements of the style approach theory of leadership (McCall et al., 1988). Behaviorally-based leadership styles include democratic, autocratic, and laissez-faire leadership. The style approach research stimulated the development of instruments such as the Leadership Behavior Questionnaire (LBQ). In their concern for people and concern for production, Blake and Mouton (1964) developed the managerial grid, later called the leadership grid, to explain how leaders helped their organizations accomplish their purpose. Klenke (1993) believed that people become leaders by learning a leadership style or set of behaviors, by participating in leadership situations, or by seeking leadership development opportunities, rather than developing individual personality traits. Burke and Day's meta-analysis (1986) supported the behavior modeling approach.

Power-influence theory. Power influence approach theory explains leadership effectiveness in terms of the amount and types of power the leader possesses and how they exercise their power. Power is important not only in influencing subordinates, but also for influencing peers, superiors and people outside the organization (Yukl, 1989). Yukl indicated that the amount of "position power necessary for leadership effectiveness depends on the nature of the organization, task, and subordinates. Leaders who lack sufficient position power to make necessary changes, reward competent subordinates, and punish or expel chronic troublemakers will find it difficult to develop a high performing organization" (p. 256).

Contingency leadership theory. Contingency leadership theory (also called leader-match theory) is one where effective leadership is contingent upon matching a leader's style to the right setting (Fiedler, 1964). It is called a contingency theory because it

“suggests that a leader’s effectiveness depends on how well the leader’s style fits into the context” (Northouse, 1997, p. 74). Fiedler, Chemers, and Mahar (1976) developed a self-paced, instruction workbook, LEADER MATCH, which taught leaders to assess and change their leadership style. Burke and Day (1986) supported the effectiveness of leader-match training programs.

Situational leadership theory (Hersy & Blanchard, 1969) focuses on specific management situations where the leader adapts the leadership style to the demands of the different situations. Northouse (1997) described the situational approach as how leaders should adapt their styles to developmental levels of followers and to the demands of various organizational settings.

Yukl (1989) indicated that situational leadership emphasized “the importance of contextual factors such as the leader’s authority and discretion, the nature of the work performed by the leader’s unit, the attributes of subordinates, and the nature of the external environment” (p. 261). Yukl further described the research surrounding this theory in two major sub categories. One approach “treats leader behavior as a dependent variable: researchers seek to discover how the situation influences behavior and how much variation occurs in managerial behavior across different types of managerial positions. The other ... seeks to discover how the situation moderates the relationship between leader attributes or behavior and leader effectiveness” (p. 262). In essence, this theory surrounds the behavior patterns of influencing others, whether in directing, coaching, supporting, or delegating.

Path-goal theory. Path-goal theory (Evans, 1970; House, 1971) focuses on “how leaders motivate higher performance by acting in ways that influence subordinates to

believe valued outcomes can be attained by making a serious effort. Aspects of the situation such as the nature of the task, the work environment and subordinate attributes determine the optimal amount of each type of leader behavior for improving subordinate satisfaction and performance” (Yukl, 1989, p. 263). The path-goal theory describes how leaders motivate subordinates to accomplish designated goals and emphasize relationships between the leader’s style and situational variables in the organization (Northouse, 1997). A meta-analysis by Wofford and Liska (1993) showed some support for the path-goal theory, but overall the results of the meta-analysis were mixed.

Leader-member exchange theory (LMX). The key concept of the leader-member exchange theory (Graen, 1976) is that leadership is a process of interactions between leaders and followers, making the leader-member exchange the focal concept of the leadership process. Yukl (1989) described the LMX theory as the “vertical dyad linkage theory” that described how “leaders develop different exchange relationships over time with different subordinates. Some subordinates are given greater influence, autonomy, and tangible benefits in return for greater loyalty, commitment, and assistance in performing administrative duties” (p. 266). Yukl described the LMX theory as situational only in the sense that leaders treated subordinates differently depending on whether they were part of the in-group or out-group.

The leader-member exchange theory examines relationships and not behaviors or traits of leaders or followers. House and Aditya (1997) asserted that a high degree of mutual influence and obligation between superiors and subordinates resulted in positive outcomes as lower turnover, and higher subordinate performance, citizenship behavior, satisfaction, and commitment. Yukl noted that the LMX theory was more descriptive

than prescriptive, describing a typical process of role making by leaders but not specifying what pattern of downward exchange relationship was optimal for leadership effectiveness.

### **Management vs. Leadership**

Since 1977 conventional wisdom has been that leaders and managers are different types of people (House & Aditya, 1997). And, the controversy regarding the distinction between leadership and managerial behaviors continues to permeate leadership literature (House & Aditya, 1997; Kotter, 1990; Yukl, 1994).

Kotter (1990) defined “leadership” as producing change and movement, incorporating vision, building/strategizing, aligning people/communicating, and motivating/inspiring employees. According to Kotter, “management” produces order and consistency and involves planning/budgeting, organizing/staffing, and controlling/problem solving. Managers are most concerned about how things get done and leaders concerned about commitment and vision (Kotter, 1990).

Burke and Litwin (1992) also distinguished between leadership and management practices. The leadership role was defined as one of providing direction and acting as a role model. Management practices, on the other hand, was described as the routine behaviors exhibited by managers as they utilize human and material resources to enact the organizational strategy in order to achieve goals.

Kotter (1996) indicated that the distinction between leadership and management was increasingly important in organizational environments that were complex, competitive and dynamic. Katz and Kahn (1978) suggested that leadership was the “influential increment over and above the mechanical compliance with routine directives

of the organization” (p. 302). Katz and Kahn also believed that leadership involved the use of influence, while management involved the use of authority.

Rost (1993) defined management as “an authority relationship between at least one manager and one subordinate who coordinate their activities to produce and sell particular goods and/or services” (p. 145). Since Rost viewed management as a relationship, the behaviors of both the manager and the employee became important variables. Rost also believed that management coordinated activities, set goals, and made decisions about staffing, jobs, and the distribution of resources necessary to achieve performance goals.

Teal (1998) suggested that management was more than technical skills -- a set of human interactions. Teal described management as a transactional relationship that derived its power from authority, which often included coercive tactics on the part of the manager.

Yukl (1994) clarified the distinction between leadership and management with the belief that *leaders* were oriented toward innovation and managers oriented toward stability. Yukl used the term “managerial leadership” in describing the overlap in the literature between management and leadership.

House and Aditya (1997) indicated that leadership was articulating an organizational vision, introducing major organizational change, providing inspiration, and dealing with high profile aspects of the external environment. House and Aditya (1977) believed that management was the implementation of the leaders’ vision and changes introduced by leaders, and the maintenance and administration of organizational infrastructures. Obviously one can be a leader without being a manager, but it is more



difficult to conceive of managers where leadership is completely absent (Sourcie, 1994). However, commonalties do exist between the concepts of management and leadership. Both are concerned with goal accomplishment, require working with people, and involve influencing others (Northouse, 1997).

Kanter (1997) concluded that managers in today's organizations must be comfortable with change and provide a sense of clarity and direction. Collins et al. (2000) indicated that organizations must embrace change as a core value. For example, the top level of the organization must think strategically to compete globally, continuously restructure the organization to meet organizational challenges, to accommodate demographic changes in the work force, meet customer demands, or embrace rapid technological changes (Gibler, Carter, & Goldsmith, 2000; Hooijberg, Hunt, & Dodge, 1997). Leaders incorporate vision and function by strategizing while managers are oriented toward stability. Therefore, this research used the term "managerial leadership development", not managerial training, to blend the controversy in the literature regarding managers and leaders.

### **Leadership Development Literature**

Leadership research is extensive, but in comparison the literature on managerial leadership development and its impact is minuscule. Not only is the managerial leadership development literature sparse, but also most studies are not empirical (Goldstein, 1980).

Only one chapter of Bass and Stogdill's (1990) *Handbook of Leadership*, and two short chapters of Gardner's (1990) *On Leadership* are devoted to leadership development. Other works (Clark & Clark, 1994; Clark, et al., 1992; Conger & Benjamin, 1999);

Gibler et al., 2000; McCall, 1998; McCall, Lombardo, & Morrison, 1988; McCauley et al., 1998; Northouse, 1997; Ulrich, Zenger, & Smallwood, 1999) focus on examples of managerial leadership development interventions from either a theoretical or a practitioner's perspective without providing major insight into managerial leadership development itself.

McCauley et al. (1998) defined managerial leadership development as “the expansion of a person's capacity to be effective in leadership roles and processes” (p. 4). Said differently by Brungardt (1996), managerial leadership development is defined as “every form of growth or stage of development in the life-cycle that promotes, encourages and assists the expansion of knowledge and expertise required to optimize one's leadership potential and performance” (p. 83). Lynham (2000) indicated that managerial leadership development involves a person's career throughout their lifetime, and was often confused with leadership education, a short-term developmental event. In addition, Klenke (1993) believed that the distinction between leadership education, leadership training, and managerial leadership development was often blurred.

Managerial leadership development literature is embedded in psychology and management, and focuses on general approaches to leadership through traits, behavioral, situational, power-influence and transformational (or charismatic) theories (Northouse, 1997). Managerial leadership development literature also includes research related to early childhood and adolescent development, the role of formal education, on-the job experiences, and specialized leadership education (Lynham 2000). Managerial leadership development, according to Conger (1992), may be grouped into four categories: leadership training through personal growth; leadership development through conceptual

understanding; leadership development through feedback; and leadership development through skill building that focuses on key leadership competencies that can be taught.

Noteworthy scholarly work has been done by Brungardt (1996) providing information on managerial leadership development as a body of knowledge and offering insights into leadership learning and training, and by Lynham (2000) in the identification of the core knowns about managerial leadership development. Lynham identified eight core knowns about managerial leadership development:

- 1) It occurs from early childhood and adolescent development.
- 2) Formal education plays a key role.
- 3) On-the-job experiences are important.
- 4) It occurs through specialized leadership education programs.
- 5) Leadership training focuses on three areas: improving a leader's knowledge, skills, and attitudes; training in success and effectiveness as a leader; and, training and education on leadership styles (Bass, 1990).
- 6) There are a number of factors that can act as potential barriers to its effectiveness.
- 7) It is a life-long process.
- 8) It is often confused with management development (Kotter, 1990; Kouzes & Posner, 1996; Yukl, 1989).

According to McCall (1998), three assertions can be made about managerial leadership development. First, challenging experiences are the primary vehicle for development. Second, the experiences that are most important are a function of the business strategy and organizational values. Third, the people who should get the

experiences are those who are best able to learn from them. However, no universal theory of managerial leadership development has emerged (Lynham, 2000).

In *Building Leaders*, Conger and Benjamin (1999) provided an in-depth assessment of the current state of managerial leadership development in corporate America and outlined the proven methods for building leadership talent at all levels. Conger and Benjamin presented three principal approaches to leadership education: individual skill development, instilling organizational values that promote leadership, and strategic interventions that promote dialogue and effect change throughout an organization. They expressed a concern that “many organizations teach and develop leadership skills that may be outdated by the time younger generations reach the senior ranks ...(and) most organizations treat leadership development casually” (p. xiii). Conger and Benjamin suggested, “tomorrow’s attributes are what we must begin to train and develop today” (p. xvi).

McCauley et al. (1998) in their *Handbook of Leadership Development* presented a model of managerial leadership development that incorporated assessment, challenge, and support in making the developmental experience more meaningful. They expanded the definition of leadership experiences to include a full range of experiences such as on-the-job experiences, developmental relationships 360-degree feedback, feedback-intensive programs, and skill-based training.

McCall (1998) presented a strategic framework for identifying and developing future executives. In addition, McCall believed that real leaders of the future are those who have the ability to learn from their experiences and remain open to continuous learning. Actually, “what’s at stake is continuous improvement” to achieve greater

performance (Kline & Saunders, 1993, p. 33). McCall provided examples of how to create an environment that supports development of talent and links business strategy with the kinds of experiences people need if they are to lead a company in fulfilling its mission. McCall's philosophy was that "leadership ability can be learned, that creating a context that supports the development of talent can become a source of competitive advantage, and that the development of leaders is itself a leadership responsibility ... Further, the primary classroom for the development of leadership skills is on-the job experience" (p. xii).

Development through job experiences, such as on-the-job-training, job-performance evaluations and feedback programs, participation in special projects or task forces, coaching or mentoring, job rotation, succession planning, and career planning have emerged as a powerful source of learning for managers (Johnson, McLaughlin, Saari, & Zimmerle, 1988; McCauley & Brutus, 1998). A common feature of developmental jobs is the "degree of challenge they offer managers" (Brutus et al., 2000, p. 368). McCauley, Ruderman, Ohlott, and Morrow (1994) asserted that challenging experiences stimulate development by providing managers with the opportunity to learn and acting as a motivator for learning.

McCall et al. (1988) believed that experiences on the job provide the skills for successful business leaders to maximize their skill development potential. They confirmed that on-the-job development, while widely recognized as important, had not received the kind of research attention that allowed practitioners to understand the magnitude of the experience. On-the-job experiences (OJT) forces "managers to learn new skills on the run, learn to act when stakes (are) high, learn to work with trying people

under trying circumstances, and learn to cope with an exhausting workload” (McCall et al., 1988, p. 59). In essence, McCall et al. believed that “developing leadership ultimately boils down to what a person does with his or her opportunities and abilities” (p. 121).

Jacobs and Jones (1995) provided a comprehensive guide to the design, delivery, and evaluation of structured on-the-job training. They contended that structured OJT as opposed to unstructured OJT experiences make a critical difference in effectiveness and efficiency. Jacobs and Jones claimed that the structured OJT system must be evaluated objectively in terms of its contribution to an organization’s performance and goals. Jacob and Jones described the five steps of their system as: prepare the trainee, present training, require responses, provide feedback, and evaluate performance. They described structured OJT as a key component within a culture where employees are encouraged to engage in continual learning activities, but not at the expense of forgetting that learning and doing go hand in hand. Therefore, Jacobs and Jones committed to the integration of learning and work as being very powerful and visionary.

The use of 360-degree feedback and multi-rater assessments are becoming key components of managerial leadership development (Antonioni, 1996; Bracken, 1994; Church & Bracken, 1997). Kuchinke (2000) expressed that feedback was a key component of any learning process. However, Kuchinke claimed that while many organizations were deliberately structuring feedback processes, they were not considering the full scope of desired outcomes or building systems to support their results. Feedback improves performance because it can result in increased self-awareness and more dialogue between leaders and subordinates.

Global leadership development focuses primarily on the study of expatriates -- their selection, training, and adjustment to foreign countries (Briscoe, 1995; Dowling, Welch, & Schuler, 1999) -- and cross-cultural training issues (Hofstede, 1986). Osman-Gani (2000) indicated that expatriate training was more prevalent in European and Japanese multinationals than with U. S. companies. Recent research has identified a broad spectrum of characteristics important for global leaders (Black, Morrison, & Gregersen, 1999; Brake, 1997; Tung & Miller, 1990; Yeung & Ready, 1995). The dimensions tend to be similar to those in U.S. research, but they are made much more complex by the nature of the global organization (Maznevski & DiStefano, 2000).

### **Performance Improvement Literature**

In describing performance, Holton (1999) distinguished between “performance” and “performance drivers.” Performance was defined as the actual outcomes produced by the organizational efforts. Performance drivers were those aspects of performance that were expected to sustain or increase system, sub-system, process or individual ability and capacity to be more effective or efficient in the future. Performance drivers and performance outcomes together portray the cause and effect relationship that exists in an organization’s strategy (Kaplan & Norton, 1993). Organization performance is directly related to performance drivers.

Burke and Litwin (1992) defined performance as “the outcome or result as well as the indicator of effort and achievement” (p. 533). These outcomes included productivity, profit, service quality and customer or employee satisfaction. Burke and Litwin believed that in a systems perspective, it was the convergence of the effects of all organizational variables that leads to performance improvement. Burke and Litwin suggested that

research demonstrated that leadership affected organizational performance, and had accounted for more variance in performance than other organizational variables. Meisel and Fearon (1996) believed that effective leadership was the “new bottom-line of organizations” (p. 180).

Kaplan and Norton (1993) suggested that there were four categories of performance measures: financial, customer, internal business process, and learning and growth. They cautioned that financial measures alone were not indicative of improved performance. Some companies have carved out a market niche and have a following of loyal customers that guarantee financial viability. However, Kaplan and Norton explained that effective and efficient production of a product without a market need also brought financial gain. They suggested that viable organizations competing in today’s business environment be cognizant of the outcomes in all four segments of performance measures because they function as an information feedback and reporting system.

Beer (1980) suggested that while financial indicators were typically the criteria used to measure organizational performance, other important criteria existed. These included a compatible work environment, equitable rewards and compensation, job security, and meaningful work. Overall, the organization must be capable of providing quality of work life in order to attract, retain, motivate, and influence employees who are committed to the organizational mission, purpose and goals.

There is little confirmation in the literature of the effectiveness of managerial leadership development interventions in enhancing performance at the group or organizational levels. Knauft (1992) indicated that there is a “lack of consensus among researchers as to what constitutes effectiveness” (p. 37). Nevertheless, organizations



continue to invest in managerial leadership development (Boyett & Boyett, 1998; Brinkerhoff & Montestino, 1995; Clark & Clark, 1994; Gibler et al., 2000; McCauley, et al., 1998). However, Ulrich et al. (1999) indicated that more research must be conducted related to “how organizational capabilities and leadership competencies lead to and are connected to desired results” (p. vii).

As organizations realize how their market value increasingly relies on the knowledge and skills of their employees (McLagan, 1997), caring about their human competency base and how it is developed makes strategic sense to them. “It is the development of workplace expertise that is becoming vital for organizations to adapt to change and maintain optimal organizational performance” (Herling, 2000, p. 9). It has been suggested that by adopting some or all of the prescribed components of a learning organization, an organization’s performance should be improved (Kline & Saunders, 1993; Kuchinke, 1995; Senge, 1992). Learning is viewed as the means to long-term performance improvement (Guns, 1996). Also, Ireland and Hitt (1999) believed that the systematic efforts to produce knowledge enhanced the organization’s ability to perform more effectively.

The high-performance leadership competency model (Holton & Naquin, 2000; Rummler & Brache, 1995) merged leadership research with performance improvement theory (Brungardt, 1996; McCauley et al., 1998; Torraco, 1999). This competency model incorporated leadership theory with Rummler and Brache’s performance model to provide a set of competencies and concrete standards for developing future managerial leadership development interventions. The high-performance leadership competency model provided a definition of managerial leadership development that included

“improving performance” as an explicit outcome (Lynham, 2000). Many authors have asserted that the best way to cope with changing times is to build high-potential employees through structured managerial leadership development systems (Gibler et al., 2000) with a valid means of assessing the impact of various training programs (Berkeley, 1988; Burke & Day, 1986; Cato, 1990; Smith, 1987).

Ulrich, et al. (1999) developed *Results Based Leadership* as a guidebook that bridges the gap between leadership theory and leadership skills, focusing on producing results that can be measured and integrated into any business strategy or corporate culture. Ulrich et al. “reframed the discussion about leadership” and provided action-oriented guidelines on how to deliver results in four specific areas: results for employees, the organization, its customers, and its investors (p. xii). “Desired results must be strategic and must align with the purposes of the organization” (Ulrich et al., 1999, p. 35). Kanter (1997) suggested that a mission statement is the motivator, as it enables people to recognize the importance of work they perform. Ulrich et al. believed that effective leaders know how to “connect their attributes with results” to improve performance (p. 1). “Understanding and measurement of both the (leadership) attributes and results of effective leaders are necessary to fulfill the goal of improving leadership quality” (Ulrich et al., 1999, p. 20).

Fiedler (1996) stated “all of the reviews of leadership training stress that we know every little about the process of leadership and managerial training that contributes to organizational performance. At least one reason for this lack of knowledge is the scarcity of meaningful and rigorous research...” (p. 244). Some definitions of managerial

leadership development include statements of achieving results, but those results are only partially clarified (Bass 1990; Bennis & Nanus, 1985; Gardner, 1990; Ulrich et al., 1999).

Ulrich et al. (1999) also believed that “leaders build not only individual commitment but also organizational capability. Organizational capability refers to the processes, practices, and activities that create value for the organization” (p. 7). According to Ulrich et al. individual capability represents “the knowledge, skill, ability, and motives of each individual employee in his or her position” (p. 58).

Nevertheless, Lynham (2000) believed that more evaluative research needs to be conducted linking performance improvement and managerial leadership development. Performance improvement, and specifically performance improvement at the organizational level, is typically inferred, implied and assumed to be an outcome of leadership and managerial leadership development (Sogunro, 1997). Yet, despite the immense amount of investment in leadership training on the part of corporations and governments, there needs to be stronger evidence that such training results in a more effective leadership behavior (Burke & Day, 1986; McCauley et al., 1998).

According to Bates (1999), attributes of performance “can be difficult to quantify because of their complexity and magnitude” (p. 51). The broad focus of performance improvement for the organization also compounds the challenge of evaluation. “Large amounts of time and money are spent on organizational performance improvement efforts without measuring the results” (p. 47). Mitchell (1983) defined performance as a collection of behaviors over time, tasks, or people. Performance measures thus represent a cluster of interrelated behaviors that reflect some underlying performance domain. Deciding what to measure involves “packaging performance from a domain into

meaningful and measurable clusters that will enhance understanding and prediction and meet the goals of measurement” (Bates, 1999, p. 51).

Despite service paid to the value of organizational results, according to Bates (1999), “training effectiveness continues to be determined largely through the measurement of performance – and often only learning – at the individual level” (p. 57). According to Bates, “a deficiency in performance measurement is the tendency of researchers to use only one criterion measure” (p. 59). “The multidimensionality of performance is regularly overlooked in research practice” (p. 56). Holton (1996) believed that HRD practitioners commonly used subjective ratings that are often contaminated as to the type of performance measure.

A review of the literature indicated that HRD practitioners must align managerial leadership development initiatives with the strategic needs of the organization (Koch & McGrath, 1996; Rummler & Brache, 1995; Swanson, 1994; Swanson & Holton, 1999). Tools that forecast the economic return of training exist (Swanson & Gradous, 1988), and, in addition, Krohn (2000) provided a decision-making model to forecast the likely strategic potential of investment in a training initiative. Krohn’s model assumed that “both ROI and inter-organizational competitive advantage must be present in a training initiative that is strategic” (p. 64). If not, Krohn believed that the “organization should treat the investment as an employee benefit and manage it accordingly” (p 64).

Today’s business environment requires that HRD not only support but also assume a pivotal role in the shaping of business strategies of organizations. However, Meldrum and Atkinson (1998) found that many organizations have not been implementing management development strategies or aligning management development

programs with strategic requirements. HRD serves a strategic role by “assuring that the competence of employees meets the organization’s present performance demands” (Torraco & Swanson, 1995, p. 11). By connecting the goals of managerial leadership development interventions to the mission of the system, HRD practitioners can explain the importance of the interventions in adding value to the organization and ultimately enhancing organizational performance.

Sometimes management training programs fail because they have no connection to real life in the company (Berry, 1990; Carlisle & Henrie, 1993) or fail to add value to corporate strategy (Swanson & Holton, 1999). Traditionally management development systems are relegated to narrowly defined support roles, where individuals are trained around current job-based deficiencies or predicted knowledge and skill needs (Olian et al., 1998). Swanson and Holton believed that to have utility or payoff to the organization, the intervention must be linked with organizational goals. Collins (2001) believed that HRD should take the lead by strategically aligning managerial leadership development systems that advance and sustain the organization’s competitive position in its market.

Researchers believe that significant financial payoffs are found among companies that emphasize training and development (Huselid, 1995; Jacobs & Jones, 1995; Lam & White, 1998; Swanson, 1994; Ulrich, 1997), and that organizations with stronger learning environments demonstrate greater organizational effectiveness (Tannenbaum, 1997). An organization’s human resources are recognized as the hidden forces behind growth, profits, and lasting value (Herling, 2000; Lau & May, 1998; Pfeiffer, 1994; Reichheld, 1996; Torraco & Swanson, 1995).

## **Outcomes of Managerial Leadership Development**

This research focused on the “outcomes” of managerial leadership development interventions. Outcomes were defined broadly to relate to the results of changes in leadership style in top management, worker satisfaction, work teams, or organizational change (Lynham & Swanson, 1997). As stated previously, there appeared to be little evidence in the literature of managerial leadership development interventions and managerial training contributing to organizational performance (Fiedler, 1996). A review of related literature on managerial leadership development did not adequately address outcomes at either the learning or performance level (Lynham, 2000). The literature indicated that the outcome of most training was to teach or improve managerial skills and improve on-the-job performance (Bass, 1990; Burke & Day, 1986; Goldstein, 1980).

Holton (1999) defined outcomes as “the measurements of effectiveness or efficiency (of the organization) relative to core outputs of the system, subsystem, process, or individual” (p. 33). Driggs (1999) emphasized that outcomes can be encapsulated as the awareness of the importance of organizational motivation and understanding, the flexibility to adapt to individual organizational needs, the openness to encourage continuing discussion and interchange, and a readiness to continue learning. Examples of outcomes in the current literature were improved subordinate and human relationships, improved knowledge skills and attitudes, improved trainee leadership and group effectiveness, improved decision-making style, sensitization of trainees to their management role, and development of a shared personal and organizational vision (Bass, 1990; Clark & Clark, 1994; McCauley et al., 1998).

To assess performance-level results requires that mission-related performance outcomes be carefully specified and connected to the mission of the system (Swanson & Holton, 1999). Swanson and Holton defined system outcomes as “units of mission-related outputs in form of goods/services that have value to the customer and that are related to the core organizational, work process, and group/individual contributors in the organization” (p. 14). The performance outcomes can be at the whole system level (organization), work process within the system (subsystems), at the work group (team) or the individual level. According to Swanson and Holton, “every intervention should lead to system outcomes at some point” (p. 69).

Senge (1990) claimed that systems thinking shifts away from a myopic view of behavior and learning to a focus on the whole and the interrelatedness of the parts, and the influence that one part has on the other components. It connects individuals, teams or groups, and organizations. Recognizing system outcomes leads to the realization that decisions, behaviors, and activities of individual employees have an effect on all the interrelated components. According to Senge, a system approach allows organizational members to see the complete organization and the influential sphere of decisions and behaviors.

Collins (2001) analyzed 54 studies on managerial leadership development from 1982-2000 and found that organizations had begun to take a more systemic approach to leadership development. According to Collins, strategic leadership and team management skills were the most distinctive changes in the content of managerial leadership development programs since Burke and Day’s (1986) study. One third of the studies in Collins’ analysis focused on organizational performance as the outcome of the

managerial leadership development experience as compared to three percent of the studies in Burke and Day's meta-analysis. Collins found formal training continues to be the primary type of managerial leadership development intervention, while job assignments was a close second.

Swanson and Holton (1999) indicated that outcomes are the driving force and energy behind the design and implementation of all managerial leadership development interventions. Outcomes of managerial leadership development interventions have the potential to affect the individual, group, and organizational performance levels (Rummler & Brache, 1995). The individual outcomes of knowledge acquisition and self-awareness building are developed over a shorter period of time or in single events, whereas skill development and behavior change are developed over longer periods of time or through multiple events (McCauley et al., 1998). The improvement of skills or changes in behaviors of individual managers is known to improve the productivity of workgroups and to enhance the ability to meet organizational goals (McCauley et al., 1998).

### **Evaluation of Managerial Leadership Development**

Today's managers recognize the impact that evaluation has on the organization, but they rarely think of evaluation as an essential part of their strategy (Bassi, Benson, & Cheney, 1996; Kaplan & Norton, 1993). Part of the problem is that few organizations utilize comprehensive evaluation strategies for management training and education (Carnevale & Schulz, 1990; Saari et al., 1988). Burke & Day (1986) indicated that many questions remained concerning the relative effectiveness of various training methods in improving learning or skill acquisition.



Outstanding practices in management development indicate that high-performance organizations always evaluate the impact of their managerial leadership development processes (Collins et al., 2000; Fulmer & Wagner, 1999; Rosenbaum & Keller, 1995). Swanson and Holton (1999) supported the HRD worldview model of evaluation where a needs assessment is conducted, the intervention is proposed, created and implemented, and the program is evaluated.

Holton (1996) believed that “evaluation of interventions is among the most critical issues faced by the field of HRD today” (p. 5). In recent years, increasing global competition, the economy, or merely a lack of understanding of the importance of human resources has placed intense pressure on human resource development (HRD) to demonstrate that programs contribute directly to the organization’s “bottom line” (Dionne, 1996; Gordon, 1987; Holton, 1996; Plant & Ryan, 1992). Unfortunately, many companies with extensive training budgets do not have a comprehensive approach to tying their training investment to bottom-line results (Phillips, 1997). Provo (2000) claimed that there is a need for “the development of human resource measurement systems that accurately demonstrate the value of investment in HR initiatives” (p. 77).

Some theorists suggest that “investment in training should be a business decision treated in a manner similar to any other capital investment that an organization would make” Krohn, 2000, p. 64). However, human capital differs from other capital investments because “the investments cannot be separated from the individual and, more specifically from the knowledge, skills, and abilities that the individual gains because of the investment” (p. 64).

HRD interventions should be linked to organizational mission, strategy, and goals, which produce results that are valued by the organization (Rummler & Brache, 1995; Swanson, 1994). However, Preskill (1997) contended that many HRD professionals were reluctant to change because they believe that evaluation and measurement were too difficult or too costly. Preskill also believed that evaluation as currently conceptualized, practiced, and researched, is not sufficient for answering many of the questions trainers and managers had about the effectiveness of organizations' training and development efforts. Walker and Bechet (1991) defined effectiveness as the extent to which human resources support long-term business planning and strategies.

Goldstein (1986) described training evaluation as the systematic collection of data regarding the success of training programs. Kraiger, Ford, & Salas (1993) indicated that it occurs when specified outcome measures were conceptually related to intended learning objectives. Kraiger et al. (1993) noted that evaluation is normally conducted to answer either of two questions: whether training objectives were achieved, or whether accomplishments of those objectives resulted in enhanced performance of the individual on the job. Evaluation also enables trainers to continuously monitor their programs and to identify points of intervention for program improvement (Martin & Kettner, 1997; Phillips, 1997).

According to Rummler and Brache (1995), evaluation is "the primary tool for linking individual performance to that of the organization, for the purpose of taking improvement action." Rummler and Brache further noted that it is only through evaluation that performance can be monitored, managed, and improved. "The quantification and measurement of expertise are necessary if an organization is to be able

to improve its performance and the performance of its human resources” (Herling, 2000, p. 11).

Kaplan and Norton (1993) believed that “effective measurement must be an integral part of the management process” (p. 143) and provides managers with a comprehensive framework to translate a company’s strategic objectives into a coherent set of performance measures. Kaplan and Norton’s philosophy can also be translated to performance improvement at the organizational level as higher competencies of the managers within the organizations translate to increased organizational effectiveness.

Dionne (1996) described three core knowns that have emerged from the current evaluation research. First, the fact that people learn something in a training activity does not mean they will apply the new knowledge in the organization. Second, a favorable organizational climate is important in the transfer of knowledge and its retention over the long term to prevent a relapse into old habits. Third, researchers, trainers, and managers judge training activities against their own standards, and the information they seek is not used for the same purposes.

According to Collins (2002), “the challenge is huge for HRD because, for the evaluation process to be meaningful for the organization, the assessment type and intensity depends on the objectives of the management development effort and the organization’s culture. Therefore, evaluation methods must be specific but also broad enough to satisfy the evaluation needs of all organizations while providing methods to conduct empirical research on outcomes of management development programs”.

Kirkpatrick’s four-level evaluation model has been used in evaluating leadership programs for forty years, and remains the most influential and prevalent evaluation

approach among practitioners (Kirkpatrick, 1996). “The power of Kirkpatrick’s model is its simplicity...” (Alliger & Janak, 1989, p. 331). The model’s simplicity is appealing to practitioners, but is also a liability. According to Holton (1996), “a shortcoming of the Kirkpatrick model is that it does not fully identify all constructs underlying the phenomena of interest, thus making evaluation impossible (p. 6). Clement (1982) also noted that essential elements were not present. Because all elements were not present in the four-level model, it was questionable whether the model could be applied universally. “The missing elements and relationships prohibit making accurate statements about system states, developing propositions, and hypotheses, and making predictions” (Holton, 1996, p. 7). Many researchers believe that the four-level evaluation model has failed the profession (Alliger & Janak, 1989; Bassi, Benson, & Cheney, 1996; Dixon, 1990; Holton, 1996; Moller & Mallin, 1996; Newstrom, 1995; Swanson 1998; Swanson & Holton, 1999) primarily because of its emphasis on reactions versus the fundamental performance results of the host organization.

Research reveals that the most commonly collected training criteria are trainee reactions (Bassi et al., 1996; Saari et al., 1988). However, Ruona, Leimbach, & Holton (1999) believed strongly that the use of reaction measures fostered a narrow view of evaluation and neglected important aspects that influenced HRD outcomes of learning, performance, and organizational results. A meta-analysis by Alliger, Tannenbaum, Bennett, Traver, & Shotland (1997) found little correlation between reactions of any type and immediate learning and concluded that “reaction measures cannot be used as surrogates of other measures” (p. 353). Alliger et al. believed that the pervasiveness of measuring reactions and the lack of studies measuring the outcomes at the organizational

performance level indicated that Kirkpatrick's four-level evaluation model (1996) was not fulfilling HRD needs in determining accurate cost-benefit ratio of management development programs.

Holton (1996) proposed a model that is the foundation of the Results Assessment System where two of the outcomes – learning and individual performance -- represented individual behaviors that a HRD intervention hoped to create. According to Holton learning was classified as a behavior in the sense that individuals choose to learn or not to learn. However, “learning is primarily internal behavior whereas performance is usually a more external one” (Holton, 1996, p. 9). Also, Kraiger et al. (1993) derived a conceptually based scheme for evaluating learning outcomes where learning outcomes were based on changes in cognitive, affective, or skill capacities. Other learning models do not view learning as a multidimensional construct, but measure learning by examining the extent to which trainees acquire relevant principles, facts, or skills (Kraiger, Ford, & Salas, 1993).

To make it possible to analyze the outcomes of managerial leadership development studies from a performance, learning, or perception perspective, Swanson and Holton (1999) developed the Results Assessment System model for evaluating HRD interventions. The outcomes in the Results Assessment System (Swanson & Holton, 1999) are at the performance, learning, and perception levels. It should be noted that only the performance and learning levels pertain to the current meta-analysis.

The performance level measures mission-related outputs, which relate to either the core organizational, process or individual goals of the organization, or the financial

outputs in terms of goods or services. An expected return on investment and linkage to the mission produces organizational results.

as “the ability of an organization to generate and generalize ideas impact. Organizations The learning level of the Results Assessment System (Swanson & Holton, 1999) confirms an increased level of knowledge or expertise (that the manager has the knowledge and is able to act on that knowledge). Measuring expertise requires that the individual demonstrate his or her behavior in a real or simulated setting. Ulrich et al. (1999) defined learning differentiate learning by generating new ideas in the form of innovation” (p. 90). Ulrich, et al. further stated that “learning also means that ideas originating in one part of the organization were codified and shared throughout the other parts, thus avoiding repetition of mistakes and guaranteeing replication of successes” (p. 90). Senge (1990) suggested that leaders and managers needed to support a learning agenda. They must send the message that personal growth is respected and valued by the organization. According to Senge, a leader’s role was one of being a model for learning, and for personal mastery and growth.

Other evaluation models used in the profession as described by Phillips (1997) include Kaufman’s five-level, CIRO, CIPP, Phillips five-level ROI, and the critical outcome technique, as well as Kaufman and Watkins’ (1996) cost-sequence analysis instrument. Kaufman’s model moves evaluation beyond the organization and examines the extent to which the performance improvement program has enhanced society and the environment surrounding the organization. The CIRO and CIPP models include a context level that defines the environment in which the evaluation will be conducted. The Phillips

model requires the evaluation to be completed at all five levels to conclude that results are achieved.

The critical outcome technique by Swanson and Mattson (1997) is an “appropriate and powerful tool for validating the financial impact of certain HRD interventions” (Swanson, 1998, p. 292). However, in most organizational activities, detailed data do not exist or are not available to realistically compute a return on investment. In such cases the cost sequence analysis instrument by Kaufman & Watkins (1996) may be employed when it is important to have criteria for deciding whether an intervention is worth funding.

Collins (2002) performed a critical review of management development interventions from 1986-2000 through a Results Assessment System lens to determine common evaluation characteristics in organizational performance improvement efforts. Collins found eighteen studies with performance-level evaluations from 1986 – 2000. Collins also discovered that when evaluating organizational performance outcomes, the researcher must make a long-term commitment to observe the pay-off for the management development intervention because organizational changes often take many years to become evident. Collins believed that organizational change often involved a change in culture, which typically occurred naturally over a long period of time as new beliefs and values were instilled in the organization. In addition, Collins indicated that the ideal assessment for organizational performance incorporated both quantitative and qualitative measurement methods. Both qualitative and quantitative methods were needed to learn more about the nature and application of transformational leadership, including how it was developed and used to create cultures in which both people and

performance were valued. Job assignments were the most common management development interventions in Collins' analysis. Collins found that results on evaluations of the effectiveness of the managerial leadership development interventions continued to be mixed. Three studies were found that had overall negative findings (Facteau et al., 1998; Moxnes & Eilertsen, 1991; Thoms & Klein, 1994).

Facteau et al. (1998) analyzed the reactions of leaders to 360-degree feedback and found that none of the perceived usefulness of peer feedback was significant. Learning at the knowledge level (Swanson & Holton, 1999) was the intended outcome of the study.

Moxnes and Eilertsen (1991) evaluated the influence of three management-training programs upon organizational climate. They found that the most process-oriented training programs changed organizational climate, as perceived by supervisors, but in an apparently negative direction, especially as far as interpersonal skills and supervisorial skills were concerned. System-level performance (Swanson & Holton, 1999) was the intended outcome variable of Moxnes and Eilertsen's study.

Thoms and Klein (1994) found that no differences were observed between participation and control groups with respect to reactions to program, learning or application of program material to on-the-job experience. Learning at the expertise level of the Results Assessment System (Swanson & Holton, 1999) was the outcome variable of Thoms and Klein's study.

The best practice in evaluation processes is to test and analyze relationship and linkages between performance measures at different levels of analysis (Collins, 2002; Swanson & Holton, 1999). For instance, how do outcomes at the process, critical subsystem and individual levels affect important mission-level outcomes? Performance-



level evaluation is normally considered to be at the system level, but a system outcome can be in sub-systems such as a functional unit, work group or team within the organization (Swanson & Holton, 1999). Collins (2002) discovered only eleven percent of managerial leadership development interventions intentionally evaluated the experience at both the performance and learning levels (Avolio & Howell, 1992; Barling, Weber, & Kelloway, 1996).

Campbell, Dunnette, Lawler, & Weick (1970) reported that HRD has been criticized for its failure to evaluate training effectiveness at the organizational level. In a meta-analysis of relationships among training criteria, Alliger and others (1997) found only three studies that collected performance measures at the organizational level. Bassi et al. (1996) discovered that of actual evaluations of HRD programs in the field that 94% of the organizations surveyed collected participant reaction information, 34% measured learning, 13% evaluated transfer of learning, and 3% measured financial impacts from training.

Collins (2002) believed that HRD should take the lead in combining evaluation theory with performance-based management development theory to create the appropriate system for measurement of organizational level performance improvement. Many studies utilize the measurement of individual learning outcomes to reflect performance at the organizational level, but there is little research that explicitly justifies that learning at the individual level translates to organizational performance (Bates, 1999). Collins also believed that HRD professionals must change from the Kirkpatrick (1996) reaction paradigm of evaluation to one that measures the effectiveness of strategic development initiatives. In addition, Collins pointed out that HRD should create and use

a model that has the appropriate constructs, which can be used as a diagnostic tool for determining the critical strategic influences to be measured along with outcomes.

Evaluation research conducted by Moller and Mallin (1996) revealed a mixture of positive and negative results. Of the respondents to a poll, most reported that they were conducting one or more types of Level 1 evaluations (Kirkpatrick, 1996). Seventy-one percent of respondents used evaluation to identify gains in learning after instruction (Level 2). Level 3 evaluations, or transfer of training, were measured by only 43% of those responding, with communication with supervisors and participant feedback being the primary methods. Evaluation on effectiveness in producing anticipated end results, Level 4, was conducted by 65% of respondents, but over half of the respondents did not report a positive benefit using Level 4 measurements. Instead, many indicated that they used Level 1 instruments for Level 4 assessments. From Moller and Mallin's research one is left wondering why so many studies at Level 4 produce negative results and why practitioners use primarily Level 1 instruments to measure Level 4 performance.

The evaluation issue is complex. Plant and Ryan (1992) indicated that managers basically wanted simple and inexpensive evaluation methods and HRD was a long way from being able to determine the impact of one type of management development program on organizational performance with enough confidence to determine an accurate cost-benefit ratio. Dionne (1996) believed that a significant difficulty in our ability to assess the impact of training "may be the lack of a unifying model and theory of research" (p. 280). More evaluative research needs to be conducted (Burke & Day, 1986; Wexley, 1984), particularly on linking managerial leadership development to performance improvement (Lynham, 2000).

## **Conclusion**

Overall there is a deficiency of scholarly knowledge about managerial leadership development, what it is, and its impact on the organization (Lynham, 2000). This literature review confirmed that more research is needed to fully understand managerial leadership development.

More empirical research is needed to link managerial leadership development with performance improvement (Lynham, 2000). The question of the impact of leadership on organizational performance is clearly a void in the current literature (Bass, 1990; Meindl & Ehrlich, 1987; Ulrich, et al., 1999). Further research is necessary “to determine when and what leadership activities influence organizational performance and the conditions under which these activities may have maximum effect” (Svyantek & DeShon, 1992). Given the increasing investment in managerial leadership development by not only American organizations but organizations worldwide, this is a particularly worrying void in the field (Lynham, 2000).

The literature also showed that the utilization of a standardized assessment instrument is important in performing evaluative studies that will adequately measure performance at the group or organizational level. Kirkpatrick’s evaluation model clearly did not contain the elements necessary to measure the impact of strategic leadership on organizational performance (Holton, 1996).

We must understand the changing and complex, overlapping skills needed for competent leadership requirements of future organizations. Therefore, strategic leadership competencies should be incorporated in future evaluative studies so as to

understand what is known about the field across multiple organizations (Collins et al., 2000).

Leadership in groups or work teams is one of the most popular and rapidly growing areas of leadership (Daugherty & Williams, 1997; Penwell, 1992; Quast & Hazucha, 1992; Riechmann, 1992). Organizational restructuring has shifted decision-making powers downward from the traditional hierarchy to more self-managed teams, empowering them in new ways. In some situations, the leadership role even rotates among members. This literature review revealed that teams or work groups are the way of the future in organizations and that more research is needed in this area.

The current literature on managerial leadership development is sparse. A review of the literature confirmed that what is really known about developmental processes aimed at growing and developing leadership capabilities and expertise, and the link between managerial leadership development and performance, still remains largely unanswered (Lynham, 2000).

### **CHAPTER 3: METHODOLOGY**

Meta-analysis was used in this research as a statistical technique to identify, aggregate, and summarize the findings of managerial leadership development studies from 1982-2001. This meta-analytic technique was used to “integrate findings across studies to control chance and other artifacts and provide a foundation for conclusions” (Hunter & Schmidt, 1990, p. 39). Glass (1976) referred to meta-analysis as “the analysis of analyses” or the “statistical analysis of a large collection of analysis results from individual studies...” (p. 3).

Meta-analysis provides a systematic approach to selecting and integrating research from studies measuring the same phenomenon. Six steps of meta-analysis are discussed in detail in this chapter within the context of explaining how the meta-analytic research for this study was conducted (Durlak & Lipsey, 1991). The six steps used in this meta-analysis were:

- 1) Identifying a common conceptual topic shared among studies in a research domain.
- 2) Operationally defining the characteristics under which studies were included and excluded.
- 3) Systematically searching the literature base for common studies.
- 4) Identifying important study characteristics that influenced outcomes and developing a coding scheme that captured these characteristics.
- 5) Extracting and analyzing comparable statistical information from research studies.
- 6) Reporting the findings in a way that accurately summarized the literature.

### **Identification of Research Topic**

Cooper (1984) set three criteria for selecting a meta-analytic topic. First, there must be a sizeable body of literature from which to draw. In a research area with only two or three comparable studies, there is little need to integrate the data. In these instances, a narrative review of the material perhaps would be the more appropriate approach. Secondly, the research should be empirically based. Meta-analysis is primarily a quantification procedure for research findings that relies on statistical findings. The third criterion for a meta-analytic topic is that studies in the area of interest should show mixed results. If all of the studies show the same results, a meta-analysis of the material would serve little purpose. Other reasons for undertaking a meta-analysis are the practical importance of the topic and the need to summarize past research before embarking on new research.

The research topic chosen for this meta-analysis was the effectiveness of managerial leadership development programs. The principal empirical support in the research literature for the effectiveness of managerial leadership development programs is Burke and Day's (1986) meta-analysis as described extensively in Chapter 2. Burke and Day determined the types of management training that were effective, to what degree they were effective, and the relative effectiveness of the different training methods in improving learning or the acquisition of skills. Burke and Day's conclusion, based on 70 studies, was "that managerial training is, on the average, moderately effective" (p. 232), and that more empirical research was needed before conclusive statements could be made. In addition, Lynham (2000) believed that the field of leadership development "may be well served by further purposeful and scholarly inquiry and study" (p. 1).

### **Criteria for Inclusion in Study**

The criteria for inclusion of studies in this research sample were:

- 1) The study was operationally defined as an organizational managerial leadership development study.
- 2) The study incorporated an intervention that involved managers, leaders, executives, officers, supervisors, and/or foremen, defined as a deliberately planned effort by an individual, group, or organization with the specific intent to enhance managerial leadership potential at the individual, group or team, or organizational level.
- 3) The study reported quantitative analyses from one of four research designs: posttest only control group (POWC); pretest-posttest with control group (PPWC); single group pretest-posttest (SGPP); and correlational studies (CORR).
- 4) The study described the treatment and outcome measures.
- 5) The study reported the group means and standard deviations, Cohen's *d*, probability level, *t*-value, Pearson's *r*, or raw data from which an effect size was determined, or the author provided this information when contacted.
- 6) The study was published in English from January 1982 through December 2001, and did not duplicate any studies that were used in Burke and Day's (1986) meta-analysis.

The research sample in this meta-analysis was a data set of managerial leadership development studies, as opposed to most research where the responses of individual subjects comprise the study sample. The information in this meta-analysis was that which

other researchers had collected and reported. Actually, the findings in the studies included in this research exhibited similar constructs and relationships and were configured in similar statistical forms (Lipsey & Wilson, 2001).

### **Identification and Retrieval of Studies**

The literature search is as important to the validity of a meta-analysis as the care that goes into drawing a random sample is to primary research. The process used to search the literature on the effectiveness of managerial leadership development from 1982-2001 involved three steps: computerized search of various databases, manual search of existing literature, and communication with subject matter experts to locate unpublished studies.

#### **Computerized Search of Databases**

A computerized search for managerial leadership development studies was conducted using WebSPIRS and Ingenta (UNCOVER) to search three databases: ERIC, PsychInfo, and Dissertation Abstracts International. This search used effectiveness, impact, influence, outcomes, and results as key words to intersect with key subject areas. Those subject areas were executive development, executive training, leadership development, leadership education, leadership training, management development, management education, management skills, management training, managerial training, supervisory training, supervisory development, 360-degree feedback, multisource feedback, multi-rater feedback, mentoring, coaching, and dyadic relationships. In addition, a computer search was conducted of five Web sites: <http://cls.binghamton.edu/library.htm>, <http://www.ari.army.mil>, <http://management.bu.edu/research/edrt/index.asp>, [www.grcl.com](http://www.grcl.com), and [leadership.center @boeing.com](http://leadership.center@boeing.com).



### **Manual Search of Literature**

Four steps were used in performing a manual search for managerial leadership development studies:

- 1) Reference lists of all studies located through the computerized search were reviewed.
- 2) All volumes of the following journals from 1982-2001 located at the Louisiana State University library were searched article-by-article: *Journal of Applied Psychology*, *Academy of Management Journal*, *Personnel Psychology*, *Group and Organization Studies/Group and Organization Management*, *Organizational Behavior and Human Decision Processes/Organizational Behavior and Human Performance*, *Human Relations*, and *Journal of Vocational Behavior*.
- 3) The table of contents was reviewed for all volumes of the following journals from 1982-2001: *Leadership Quarterly*, *The Journal of Leadership Studies*, *Journal of Management Development*, *Organizational Dynamics*, *Human Resource Development Quarterly*, and *Human Resource Management*.
- 4) All studies cited in *The Impact of Leadership* by Clark, Clark, & Campbell (1992) were reviewed.

### **Search for Unpublished Studies**

A meta-analysis is not considered complete if a subset of the population is intentionally omitted. To omit dissertations or other unpublished research would assume that the direction and magnitude of effect is the same in published and unpublished works. Efforts were taken to prevent the “file drawer problem” where “journals are filled

with five percent of the studies that show Type I error, while the file drawers back at the lab are filled with 95% of the studies that show non-significant results” (Rosenthal, 1984, p. 107). A search for unpublished manuscripts was conducted to help ensure that findings from this meta-analysis were not biased due to the absence of unobserved and unobservable effect sizes (Lipsey & Wilson, 2001).

Unpublished studies, and additional studies not found through the computerized and manual literature search processes, were sought through contacts with various leaders in the management development field:

- 1) E-mail was sent to all senior authors of articles located through the computerized search and to the second and third authors when they were from an institution different from the senior author. A copy of the e-mail can be found in Appendix A.
- 2) Contacts were made with individuals at the Center for Creative Leadership who were likely to have knowledge about available managerial leadership development studies.
- 3) Presenters on leadership or management development at conferences of the *Academy of Human Resource Development* from 1998-2001, and the *Society of Industrial and Organizational Psychology* in 2000 were contacted.

Although an effort was made to locate all published and unpublished managerial leadership development studies from 1982-2001 through these search processes, there is no claim that every such study was found.

### **Sorting and Merging of the Pertinent Literature**

The sources and abstracts for citations uncovered through the computerized searches were printed, read, and evaluated against the criteria established for inclusion in the study. In situations where it was difficult to determine if the study met the criteria, the abstract was retained in the pool of potential studies. Often it was difficult to determine from an abstract if the participants were appropriate for this meta-analysis, or if the appropriate statistical analysis appeared in the study.

Abstracts that either reported statistical findings or cited specific managerial leadership development interventions with terms such as “outcome”, “finding”, or “research” were retained as potential sources. Abstracts that described philosophical or theoretical issues, or reviewed an existing program that did not report intervention outcomes or statistical information, were generally not included. This decision was made after an investigation of approximately thirty sources and confirming that these articles were not empirically based. Because of the large number of duplications in the reported citations, the abstracts were compared and discarded when the abstract was a cite of the same article. Therefore, only one abstract was retained when duplicates were found.

A hard copy of the full article of potential studies was obtained through interlibrary borrowing, photocopied from existing journals in the Louisiana State University Library, or ordered from University Microfilm Incorporated. Studies were placed in a manila folder, with a label indicating the name of the author(s) and the title of the article. The folders were filed alphabetically.

Each of the articles were read, analyzed, and retained in the active sample pool if they met the criteria for inclusion in this meta-analysis. Duplicate publications were

found and discarded. From the literature search and reduction process, a total of 83 studies focusing on formal training interventions became the final sample used in this meta-analysis. A list of the meta-analysis sample can be found in Appendix B.

### **Coding Research Findings and Study Characteristics**

Of key interest to this research was the relationship between the intervention outcome and other key features of the study, specifically the intervention type and research design. Three different steps were followed in coding the managerial leadership development studies in this meta-analysis: development of a coding form, including the determination of key study characteristics and their definitions; development of coding instructions; and determination of coder reliability.

#### **Coding Form**

Coding forms are information-gathering instruments used by the researcher to identify pertinent information from the study that is of importance in the meta-analysis. Coding forms can be equated with questionnaires or interview forms in other types of research. Part of the work of this meta-analysis was to ask the “right” questions and probe the study in the same manner as an interviewer might interview an individual or construct a questionnaire to elicit responses (Durlak & Lipsey, 1991).

A coding form was developed to capture study-specific items such as the author’s name, publication type, and year of publication. Additional information coded was job classification level, organization type, country where program occurred, program name, sample size, intervention type, content focus, outcome category, outcome variables measured, measurement instrument, method of measurement, and applicable statistical data.

Two coding forms were initially developed: one for control-comparison groups and pretest-posttest studies, and one for correlational studies. The forms were color coded to permit easy entry of data into a computer file. The coding forms were condensed into one form following pilot testing with 20 studies and a decision to include two additional study characteristics: Cronbach's alpha and the expected direction of the outcome. The researcher returned to the previously coded studies to ensure that coding was updated with the change of the coding form. The final version of the coding form was used to classify all studies in this meta-analysis. Appendix C contains the final revision of the coding form.

The key focus of the coding form was the outcome category of each managerial leadership development study. A separate coding form was completed for each outcome relationship in a study, and thereby causing a study with multiple outcomes to have more than one coding form. This process of using separate coding forms for each outcome facilitated data entry and possibly reduced errors.

Some studies in the sample incorporated more than one independent treatment group that lent themselves to more than one comparison in the study per outcome category. Multiple comparisons also occurred when multiple measures were made based upon multiple time intervals following the intervention. In these studies, the first reported measurement between the groups following the intervention was used as the comparison of interest in this meta-analysis, thus eliminating multiple studies based on time intervals or time series designs. The most simple study in the meta-analysis provided a single comparison such as a pretest-posttest result for a single group; however, studies that

yielded multiple independent comparisons were coded separately so that they could later be aggregated by effect sizes according to outcome category and research design.

### **Coding Information**

The coding form requested 18 pieces of information from each study. This information was organized into four categories: study identification, participant characteristics, intervention characteristics, and outcome characteristics. A brief explanation of the categories on the coding form used in this meta-analysis is explained below by category of information. It is most useful in the explanation that follows to refer to Appendix C as it provides the details for the following discussion.

Study Identification. The study identification category contained information that served two functions. The first was simply a procedural function constructed to organize a large number of studies. The study identification number, the author's last name, and the publication year served this function. The second function was to collect information for use in this analysis regarding publishing patterns.

Participant Characteristics. Information about the participants who participated in the managerial leadership development interventions was collected from each study. Participant information collected was: job classification level, type of organization in which the participant worked, and the country in which the program occurred. The job classification level was designated as entry level (supervisor/foreman), mid-manager, and top management levels. Some studies had mixed levels of participants. Others specified leaders as the participants of the intervention. In those circumstances they were documented under mixed and other respectively.

The organization type was coded for each study. When a study did not report the organization type or provided an organization type that was not anticipated on the coding form, the term other was coded accordingly.

The country where the program occurred was recorded as United States or non-United States. If the company was listed as multinational and did not specify the exact location where the intervention occurred, both were checked on the coding form. In addition, some studies did not report the organization type and unknown served as the appropriate code.

Intervention Characteristics. The specific name of the managerial leadership development program and the size of the participant group were documented on the coding form. The intervention type and the content area were also captured as key study variables in this meta-analysis. In-depth descriptions of these two variables will be provided in separate sections later in this chapter, as they are pertinent to the objectives of this research.

Outcome Characteristics and Related Statistical Information. The coding form allowed for eight outcome variables to be identified in the meta-analysis sample: knowledge-objective; knowledge-subjective; expertise-objective; expertise-subjective; financial-objective; financial-subjective; system-objective; and system-subjective. Definitions of these eight outcome categories can be found in a later section of this chapter. Also captured on the coding form was the method of measurement: self-perception, other person, other objective instrument, or company records.

The statistical information used for calculating effect size in this meta-analysis (either means and standard deviation, Cohen's d, Pearson's r, F or t-value, or p levels),

was documented on the coding form. Typically, only part of the statistical information was contained in any one study. For instance, a study sometimes contained a t-test statistic, but not the standard deviations associated with the control and experimental groups. This was not a problem because effect sizes were obtained through algebraic conversions of a t-test statistic to a Cohen's d and a Pearson's r (Hunter & Schmidt, 1990, p. 272). When sufficient statistical information was not provided, the researcher e-mailed the author attempting to obtain the appropriate information. If a study did not report statistical information necessary for the meta-analysis, it was excluded because it did not meet the previously established criteria for inclusion in the study. Captured on the coding form also was the size of the control and experimental samples, and the direction of effect (whether the result supported or did not support the hypothesis under investigation).

### **Coding Instructions**

Detailed coding instructions were developed during the pilot phase of the coding process to ensure consistency in coding studies especially where study characteristics were ambiguous. Appendix D contains the coding instructions used to guide the coding decisions in this meta-analysis.

### **Coding Reliability**

Specific measures were developed to address the reliability of the coding procedures. First, the researcher randomly identified 20 studies and coded them twice. Ninety five percent (95%) of the studies were coded the same during the second round of coding.

Second, two dissertation committee faculty members, each with a Ph.D. degree and considerable knowledge and expertise in leadership and management development,



coded the same random sample of 20 studies. The researcher independently provided a short training session for each coding team member on the coding scheme, coding definitions, and the task prior to the coding procedure. Both members of the coding team independently coded each of the 20 studies in the coding sample. The researcher met individually with each coding team member to compare coding elements. There was an 88% agreement in ratings over all studies and variables with the first coder. With the second coding team member, there was a 92% agreement in ratings assigned. When uncertainty existed on a coding item, the coding team members discussed the uncertainty and resolved the coding based upon that discussion. The primary area of uncertainty was in the understanding of how variables were defined. Discussion of definitions clarified the inconsistency of ratings and eventually resulted in 100% agreement with both raters.

### **Key Coding Characteristics and Definitions**

The four key study characteristics in this meta-analysis are intervention type, content focus, outcome category, and research design. The intervention types were defined using McCauley, Moxley, and Van Velsor's (1998) full range of leadership development interventions. Content focus definitions were determined from the high performance leadership competency model (Holton & Naquin, 2000). The researcher defined intervention outcomes using the Results Assessment System (Swanson & Holton, 1999) and Burke and Day's (1986) meta-analysis.

#### **Intervention Types**

Intervention types for this meta-analysis were formal training, developmental relationships, on-the-job experiences, and feedback programs. In contrast, Burke and Day (1986) used intervention methods rather than intervention types: lecture, lecture/group

discussion, leader match, sensitivity training, behavioral modeling, lecture/group discussion with role-playing or practice, and multiple techniques. Intervention types were defined for this meta-analysis as follows:

- 1) Formal training programs: Structured training programs in a formal setting either in the workplace or in a location outside the organization that were designed to develop the individual employee.
- 2) Developmental relationships: Experiences where another individual influences the manager's personal development. Relationships that individuals regard as a key source of assessment, challenge, and support. Examples of these relationships included one-on-one mentoring, coaching, dyadic relationships or leader-match. Developmental relationships were those that provided needed support for the enhancement of an individual's career development and organization experience. In these relationships the parties had knowledge of one another and were ones where both parties could potentially benefit. Mentoring experiences consisted of relationships where one person invested their time, knowledge, and effort in increasing and improving another person's growth, knowledge, and skills.
- 3) On-the-job experiences: Experiences that stretched people, pushed them out of their comfort zone, and required them to think and act differently. On-the-job interventions included an entire job such as redesigning a system or a piece of job, or serving on a temporary task force. Job assignments not only included work that was "assigned", but also included an experience for which the manager volunteered. This category also included development

experiences with managers learning, growing, and undergoing personal change as a result of the roles, responsibilities, and tasks they encountered in their jobs.

- 4) Feedback programs: Assessment data fed back to participants in a safe, supportive environment and a systematic collection of opinions about a manager's performance from a wide range of co-workers, including peers, direct subordinates, the boss, the boss' peers, and people outside the organization (360-degree feedback). Feedback programs offered major opportunity for organizational members to improve the quality of their work relationships, helped appraisers and appraisees effectively define the quality requirements in their work relationships, taught individuals how to give and receive constructive feedback, and provided a structure for discussing the undiscussables (Antonioni, 1996). Assessment in formal feedback settings was included in this intervention type.

- 5) Mixed: Interventions with more than one intervention type.

### **Content Focus of Interventions**

The content focus of interventions was categorized as problem solving and decision-making; strategic stewardship; employee performance; human relations; and job and work redesign. A mixed category was used when more than one content focus category was incorporated in the managerial leadership development program. This content focus category was defined as general management.

Burke and Day (1986) incorporated the following training content areas in their analysis: general management, human relations, self-awareness, problem solving/decision

making, rater training, and motivation/values. (See Chapter 2 for more details of Burke and Day's study.) A primary difference from Burke and Day's study was the inclusion of strategic stewardship, focusing primarily at the organizational level (Rummler & Brache, 1995), and job and work redesign, focusing primarily on work teams. Content focus categories for interventions were defined in this meta-analysis as follows:

- 1) Problem-solving and decision-making: The emphasis focused on generalized problem-solving or decision-making skills that were applicable to a wide range of work problems that managers encountered.
- 2) Strategic stewardship: Participants learned about at least one of the following topics on how to develop, manage, and lead an overall organizational program or sub-unit of an organization:
  - a) Knowledge of industry trends or perspectives.
  - b) The importance of innovation.
  - c) How to transform organizational culture and values.
  - d) How to implement and institutionalize change.
  - e) How to communicate strategy and performance information to relevant subsystems.
  - f) The understanding of organizational politics.
  - g) The measurement of organizational outputs against strategic goals to promote continuous organizational improvement.
  - h) The ability to acquire and allocate adequate resources conducive to achieving high performance.

- i) How to develop organizational systems (personnel, technology and control) that enable effective performance.
- 3) Employee performance: Participants learned about at least one of the following management topics:
  - a) Continuous improvement through feedback.
  - b) Assessment of individuals' knowledge, skills, and abilities.
  - c) Writing performance objectives and performance appraisals that effectively contribute to the development of employees.
  - d) Creating nurturing relationships through coaching and mentoring that bring out the best in employees to help them reach their highest level of achievement.
  - e) Developing clear employee goals.
  - f) Using motivation to build employee commitment.
  - g) Increasing managers' motivation.
  - h) Modifying a manager's values or attitudes.
  - i) Adapting quickly to the organizational culture.
- 4) Human relations: Participants learned about at least one aspect of the following:
  - a) Interpersonal skills that enable them to supervise, resolve personnel conflicts, collaborate, and communicate effectively.
  - b) Human relations problems of leadership or supervision.
  - c) Appropriate attitudes toward employees.
  - d) Linkage of the right employee to the right job.

- e) Linkage of human resource policies and procedures to job activities to enhance organizational and individual performance.
  - f) Development of future leaders whose values closely matches the organization.
  - g) Evaluation of organizational policies and procedures in relation to employee performance.
  - h) The ability to provide an open atmosphere.
- 5) Job and work redesign: Participants learned about at least one of the following methods of how to redesign jobs that lead to employee satisfaction:
- a) Creation of meaningful jobs or the development of job characteristics that motivate employees to do their best work.
  - b) The appropriateness of forming work groups, becoming proficient in team development, and the enhancement of group dynamics to improve effectiveness in working together through confrontation and consensual resolution of conflicts.
  - c) The quality of group decisions and verbal behaviors.
- 6) Mixed (General Management): Participants learned topics in more than one of the defined content focus categories above.

### **Managerial Leadership Development Outcomes**

The most pertinent variable to this meta-analysis was the outcome result of each managerial leadership development intervention. Outcomes were defined broadly in this research to relate to the results of changes in leadership style in top management, worker satisfaction, work teams, or organizational climate (Lynham & Swanson, 1997).

All studies in this sample were coded as having either a performance- or learning-level outcome as defined by the Results Assessment System (Swanson & Holton, 1999). Performance-level outcomes were further categorized as system or financial results. Learning-level outcomes were delineated into knowledge or expertise results. It should be noted that perception outcomes were not included in this research. Swanson and Holton's definitions were used to define outcome categories as follows:

- 1) System outcomes: Units of mission-related outputs in form of goods/services, having value to the customer, that were related to the core organizational, work, process, and group or individual contributor to the organization.
- 2) Financial outcomes: Conversion of output units of goods/services attributable to the intervention into money and financial interpretation.
- 3) Expertise outcomes: Human behaviors having effective results and optimal efficiency acquired through study and experience within a specialized domain.
- 4) Knowledge outcomes: The mental achievement acquired through study and experience. Learning, in the context of this meta-analysis was a process of cognitively reorganizing and reconstructing how one understands oneself and the world, and used that understanding to guide future action (McCauley, 2000).

For their meta-analysis, Burke and Day (1986) developed four criterion-measure categories on the basis of two dimensions: level of criterion and subjectivity-objectivity. Therefore, their criterion categories were subjective learning, objective learning, subjective behavior and objective behavior. Outcome categories were redefined for this research to reflect the appropriate subjective-objective outcome. Therefore, the outcome

categories for this meta-analysis were based upon a combination of Burke and Day's model (1986) and Swanson and Holton's model (1999).

- 1) Knowledge - Subjective: Principles, facts, attitudes, and skills learned during or by the end of training as communicated in statements of opinion, belief, or judgment completed by the participant or trainer.
- 2) Knowledge - Objective: Principles, facts, attitudes, and skills learned during or by the end of training by objective means, such as number of errors made or number of solutions reached, or by standardized test.
- 3) Behavior (Expertise) - Subjective: Measures that evaluate changes in on-the-job behavior perceived by participants, or global perceptions by peers or a supervisor.
- 4) Behavior (Expertise) - Objective: Tangible results that evaluate changes in on-the-job behavior or supervisor ratings of specific observable behaviors.
- 5) Financial or System Results (Performance) - Subjective: Organization results perceived by respondents, not reported by company records, (e.g., subordinates' job satisfaction or commitment to the organization,) and group effectiveness perceived by subordinates.
- 6) Financial or System Results (Performance) - Objective: Tangible results, such as reduced costs, improved quality or quantity, promotions, and reduced number of errors in making performance ratings.

### **Study Designs**

Research studies often utilize different study designs to address the same research question. Many meta-analysts, such as Burke and Day (1986), limited their research only



to studies with control groups. However, every meta-analysis requires judgments about what data to include, and these choices ultimately impact interpretation results of the meta-analysis (Wanous, Sullivan, & Malinak, 1989). This meta-analysis was no exception.

When a meta-analyst is faced with multiple research designs, he/she has one of two choices: integrate all research design types to create one overall effect size, or conduct separate meta-analyses of the studies based upon research design types and create separate effect sizes per research design. According to Hunter and Schmidt (1990, p. 339), the data from studies with different research designs “must be analyzed in different ways using different formulas for sampling error.” To conduct individual meta-analyses by the type of research design provides an aggregation of similar studies and the ability to test the research design as a moderator variable.

In the current research, separate meta-analyses were conducted based upon the type of research design. Four research designs were initially included in this meta-analysis: posttest only control group (POWC); pretest-posttest with control group (PPWC); single group pretest-posttest (SGPP); and correlational studies (CORR). Correlational studies were later dropped from the meta-analysis because there were too few studies to conduct a meta-analysis with significance. The final three study designs are described below as they relate to the context of this meta-analysis.

Carlson and Schmidt refined Hunter and Schmidt’s (1990) effect size formulas, allowing for pretest-posttest research designs to be included in meta-analytic research. For consistency purposes, Carlson and Schmidt’s formulas were used to determine effect sizes in each of the three meta-analyses in this research (See Table 1).

### **Posttest Only With Control Design (POWC)**

The POWC design involved one or more variables measured in two or more independent groups, a control and a treatment group, with the variables compared between groups. The POWC design was unique, as it is the only design in this meta-analysis that did not contain pretest scores. The statistics typically reported in these research studies were the means and standard deviation upon which the two groups, the intervention and control (or comparison) group, were compared.

An effect size for POWC studies was determined using Carlson & Schmidt's formula (1999, p. 854) as shown in Table 1. The effect size for POWC studies in this meta-analysis was computed as the difference between the mean posttest scores of the treatment and control groups divided by the pooled standard deviation of the two groups. Carlson and Schmidt's formula for computation of the pooled standard deviation of the posttest trained and untrained groups also was used to compute the standard deviation (p.

Table 1: Formulas for Determination of Effect Sizes

Research Design	Formula for determination of effect size	Reference
<b>POWC</b>	$d = (T_2 - C_2)/S_{T_2C_2} = ES_{POWC}$ $S_{T_2C_2} = ((N_T - 1)S_{T_2}^2 + (N_C - 1)S_{C_2}^2 / N_T + N_C - 2)^{.05}$	Carlson & Schmidt, 1999 p. 852, 855
<b>PPWC</b>	$d = [(T_2 - T_1) - (C_2 - C_1)]/S_{T_1C_1} = ES_{PPWC}$ $S_{T_1C_1} = ((N_T - 1)S_{T_1}^2 + (N_C - 1)S_{C_1}^2 / N_T + N_C - 2)^{.5}$	Carlson & Schmidt, 1999 p. 852, 855
<b>SGPP</b>	$d = (T_2 - T_1)/S_{T_1} = ES_{SGPP}$	Carlson & Schmidt, 1999 p. 852

Note. T and C are the group means on dependent variables for treatment and control groups, respectively. The subscripts 1 and 2 represent pre-treatment and post-treatment, respectively. S represents a dependent variable measure's standard deviation such that  $S_{T_1C_1}$  represent the pooled standard deviation of the treatment and control group pretreatment dependent assessments. A  $d$  represents the effect size.

855). The purpose of using the pooled variance of treatment and control groups was to reduce the amount of sampling error associated with the effect size estimate. Essentially, the effect size for POWC designs was the normalized difference between a trained and untrained group.

### **Pretest-Posttest With Control Design (PPWC)**

Another research design included in this meta-analysis involved both a pretest-posttest contrast within a treatment group and a comparison of a treatment group with a control group. The pretest-posttest contrast of the PPWC design compared the mean and standard deviation on a variable that was measured at one time with the mean and standard deviation on the variable measured at a later time. Specifically, the pretest-posttest effect size reflected the amount of change within the treatment group. The comparison of an experimental and control group reflected the difference between the two groups.

The mean and standard deviation for the pretest and posttest measures in the treatment and control groups served as the primary statistical data from which an effect size was determined. Once again, as shown in Table 1, Carlson and Schmidt's (1999, p. 852) formula for determining an effect size was used with PPWC designs. The effect size was computed by comparing the difference in average gain scores between the trained and untrained groups. Specifically, the formula subtracted the raw mean difference of the control group pretest-posttest scores from the raw mean difference of the treatment pretest-posttest scores and divided the resultant average gain by the pooled standard deviation of the training and control groups pre-training dependent variable assessments. Carlson and Schmidt's formula for pooled standard deviation used the pre-training

dependent variable variance for the trained group and the pre-training dependent variable for the comparison (control) group (p. 855).

Of important note in this meta-analysis was the use of the pre-training standard deviation in calculating effect sizes in PPWC and SGPP designs. Glass, McGaw, and Smith (1981) argued for using the standard deviation of pre-training, as post-training may be altered as individual differences may interact with training methods employed, resulting in participants learning at different rates. These differences may be due to inattention or localized distractions in the learning environment, differing opportunities for participation, or differential exposure to treatment. This would also result in varying rates of learning and more variable scores on the post-training dependent variable measure, more variable than would have been observed on the pre-training assessment.

Carlson and Schmidt (1999) defined an effect size for PPWC designs as “the normalized difference in the gain scores between the trained group and the untrained comparison group” (p. 855). Effect sizes from PPWC design studies are based upon more information about training than effect sizes from other research designs because PPWC uses both a pretest-posttest measurement and control group, and uses the pre-training dependent variable standard deviation in determining the effect size. According to Carlson and Schmidt, PPWC studies are considered “the standard of accuracy because they most closely approximate current conceptualizations of desired effect size” (p. 853), as they are based on more information than any other design.

### **Single Group Pretest-Posttest Design (SGPP)**

The SGPP research design was used in studies in this meta-analysis to examine change in the intervention group. The standardized effect size for SGPP studies was

determined by comparing the pretest and posttest mean scores for the trained group and divided the resultant comparison by the standard deviation of the pre-training dependent variable measure (Hunter & Schmidt, 1990, p 369). The SGPP design was unique in this meta-analysis, as it did not include a control group.

There is a controversy in the literature as to whether SGPP studies should be included. A common conclusion among some researchers is that data from single group pretest-posttest designs “upwardly bias the mean treatment effect estimates derived from meta-analysis” (Lipsey & Wilson, 1993, p. 1194). Due to inattention or localized distractions in the learning environment, or differing opportunities for participation, not all participants in a training program receive equal amounts of treatment. This would also result in varying rates of learning and more variable scores on the post-training dependent variable measure.

Nevertheless, the SGPP design is frequently used to evaluate training programs (Carlson & Schmidt, 1999), and to measure individual growth and learning. Actually, Hunter and Schmidt (1990) demonstrated that “under most circumstances the within-subjects design is far superior to the between-subjects design” (p. 339). “If the dependent variable is measured with high reliability...(p. 340), the within-subjects design has a much higher statistical power than does the independent groups subjects design” (p. 341). In addition, they suggest that the treatment by subject interaction, (i.e., the individual differences of the participants’ and their responses to the intervention) is more easily detected in pretest-posttest research design studies, especially if the treatment effect is the same for all subjects and the dependent variable is measured with higher reliability.

### **Data Reduction Techniques**

Many of the studies had more than one test of significance relevant to the research questions. Some authors reported several quantitative measures for a single outcome category. In such cases, for example, the means and standard deviations were combined, where possible, through weighted averaging to provide one overall statistical value for that outcome category, a single aggregated measure of the effectiveness of managerial leadership development (Hunter & Schmidt, 1990).

To average individual p-values, the z-score for each p-value was obtained and assigned positive or negative sign based upon direction of results. The mean z-score was then obtained by adding individual z-scores and dividing by the number of scores added. The resulting p-value was obtained from the table of probabilities associated with observed values of z in the normal distribution.

In studies where t-value, p-value and the standard difference (d) were also reported, the researcher deferred to the means and standard deviation as the primary set of statistics from which to determine an effect size, followed by t-value when available. In correlational studies, when Pearson's  $r$  and  $r^2$  both were reported, the researcher chose the Pearson's  $r$  as the statistic to use for determining the effect size (Hunter & Schmidt, 1990).

### **Unit of Analysis**

The unit of analysis on which the effect sizes were combined for this meta-analysis was the outcome measure of the study. When more than one dependent measure was used to test the same relationship in a single study, the resulting data was weighted, where possible, to produce one effect size per outcome per study. However, as Glass,

McGaw, and Smith (1981) noted, when study effects are aggregated within each study to produce an average effect by study, valuable information about the individual differences is lost. Interaction effects, multiple main effects, and other statistical artifacts are not recoverable if they are contained in one effect size statistic representing the study. However, Glass, McGaw and Smith asserted that with a large number of studies with multiple effect sizes, the influence of any one set of effect sizes, on average, would not dramatically influence the outcome of the meta-analytic statistics.

In this meta-analysis, some studies described two separate treatment groups in relation to a comparison or control group, or, in other words, had multiple independent interventions within the same study. Hunter, Schmidt, and Jackson (1982) distinguished this approach as a fully replicated design, which represents separate studies that use the same designs and tests. This meta-analysis employed Hunter and Schmidt's recommendation that the multiple outcomes be treated as if they were independent studies. These independent interventions often had similar outcomes and were recorded accordingly, with effect sizes computed for each intervention.

In some studies, researchers used tests that were clearly better assessments of the relationship under investigation than the other instruments they employed. This was evident when a self-developed measurement was used in conjunction with a widely used standardized instrument, and the researcher failed to report reliability or validity of his/her instrument. In these circumstances, a judgment was made to use an effect size computed from the standardized instrument instead of the one developed by the researcher.

### **Software for Analysis**

All codes and study reference data from the coding forms were entered into *Comprehensive Meta-Analysis, Version 1.0.23* a software program generated by Borenstein and Rothstein (1999) of Biostat, Inc. This software was chosen because statistical procedures were adopted from Hunter and Schmidt (1990), allowed for synthesis of data from multiple studies, and provided a means for determining the source of variation when effect sizes differed significantly.

### **Effect Size Estimates**

The key variable in conducting this meta-analysis, the statistic around which the meta-analysis revolved, was the effect size. Lipsey and Wilson (2001) defined effect size as “a statistic that encoded the critical quantitative information from each relevant study finding” (p. 3).

An effect size can be expressed in many ways (Durlak & Lipsey, 1991). For studies with a correlational association (prior to discarding them from this meta-analysis), the product moment correlation or its variants was used as the metric for calculating an effect size. For studies concerned with group differences related to the effectiveness of managerial leadership development, the most common metric was the standardized difference between group means or the difference between the means divided by the pooled standard deviation of the scores within the group as previously explained. This metric was Cohen’s *d*. The effect size is an index that transforms the unique data from each reviewed study into a common metric. In an experimental versus control group study, the effect size of 1.0 reflects that the experimental group changed one standard deviation more than did the control group.



At least one effect size was derived from each study in the meta-analysis sample. The effect size index was the dependent variable in the analysis, and variables describing the nature and circumstances of the studies were independent variables presumed capable of influencing the effect size magnitude (Durlak & Lipsey, 1991).

The sign for the effect size statistic was an important element documented in this research that impacted the overall findings of the meta-analysis. The effect size sign was determined by whether the difference between groups was in the expected direction. For instance, a positive sign was assigned to the effect size when the intervention group did “better” than the control group or the posttest score was “better” than the pretest score. Likewise, a negative sign was assigned when the intervention group did “worse” than the control group (Lipsey & Wilson, 2001). Of special note were situations where high scores on the dependent variable intentionally indicated worse performance, as the signs were reversed in determining those effect sizes. Positive signs typically were given when treatment groups outperformed control groups and negative signs for the reverse outcome.

### **Interpretation of Effect Sizes**

To interpret the results of this meta-analysis, one should understand the standard employed that allows for meaningful interpretation of effect sizes. Several standards exist in the literature to assess the meaningfulness of an effect size (Cohen, 1977; Glass, McGaw, & Smith, 1981; Lipsey, 1990). Cohen suggested .2 as minimal effect, .5 as a moderate effect, and .8 as a meaningful effect. Lipsey categorized effect sizes into three groups. These groups and their range of effect sizes are small effect (less than .32), medium effect (.32-.55), and large effect (greater than .55). For purposes of this meta-

analysis, effect sizes less than .32 are interpreted as small, medium effect are .32 - .65, and greater than .65 are a significant effect.

### **Correction for Statistical Artifacts**

Before a correction could be made for error of measurement, all Cohen's  $d$  statistics were converted to Pearson's  $r$  using Hunter and Schmidt's (1990) formula (p. 272). In other words, all Cohen's  $d$  metrics were calculated as a relationship between the independent and dependent variable to reflect group differences using Hunter and Schmidt's formula for calculating Pearson's  $r$  from the Cohen's  $d$ -value. This approach was used because Hunter and Schmidt advised that "the simplest way to do a meta-analysis correcting for artifacts such as ... imperfect construct validity is to do the meta-analysis using  $r$ " (p. 284).

Using Hunter and Schmidt's (1990) approach, correction for artifacts in this meta-analysis was done in four steps (p. 284):

- 1) Convert all the Cohen's  $d$ 's to Pearson's  $r$ 's.
- 2) Use the formulas by Hunter and Schmidt to conduct the meta-analysis on Pearson's  $r$  correcting for all possible artifacts.
- 3) Convert the final results for the mean correlation to a mean effect size using the conversion formula for Pearson's  $r$  to Cohen's  $d$ .
- 4) Convert the standard deviation of correlations to the standard deviation for effect sizes.

The raw effect sizes coded in this meta-analysis were not used as the data for analysis without adjustment to correct "inadequacies the index possessed as a statistical estimator" (Durlak & Lipsey, 1991, p. 310). If all studies in this meta-analysis had been

conducted perfectly the effect size obtained would have been used directly to estimate the distribution of effect sizes across the managerial leadership development literature. However, studies are never perfect. Hunter and Schmidt (1990, p. 45) identified 11 artifacts that alter the observed effect size in comparison to the actual. Artifacts are errors that occur in the individual studies produced by study imperfections, which can potentially be corrected by using statistical information.

Three study artifacts were the focus of imperfect reliability in this meta-analysis: sampling error, error of measurement, and range restriction. The objective for adjusting for artifacts was to enable the findings to be as close as possible to estimating the magnitude of the relationship represented in an effect size, as it would appear under ideal research circumstances (Lipsey & Wilson, 2001). “Most artifacts not controlled have the effect of lowering the observed effect size” (Hunter & Schmidt, 1990, p. 291). In more simple words, effect sizes observed in studies tend to be biased toward zero because of artifacts. According to Hunter and Schmidt, effect size estimates are subject to distortion when based on measures that are unreliable, invalid, or restricted in range.

### **Sampling Error Variance**

According to Hunter and Schmidt (1990, p.44), “the most damaging artifact in narrative reviews has been sampling error.” Therefore, the first artifact corrected in this meta-analysis was the biased effect size when estimated on small samples. Since effect sizes were derived from sample statistics (means, standard deviations, correlations), their statistical properties depended in part on the underlying sample size (Lipsey & Wilson, 2001). Because the sample size varied, a weighted effect size was calculated so that studies of different sizes were not treated as though they made the same contribution to

the conclusions. To clarify, larger samples were weighted more in statistical computations than those based on smaller sample sizes using Hunter and Schmidt's techniques.

The computer software, Comprehensive Meta-Analysis, automatically adjusted for sample size bias using Hunter and Schmidt techniques to produce a "bare bones" meta-analysis. A "bare bones" meta-analysis of determining a mean effect size for all studies in the sample corrects for sampling error but makes no correction to the mean or variance for artifacts such as error of measurement or range restriction.

### **Error of Measurement**

Error of measurement is the random measurement error assessed as unreliability of the measure (Hunter & Schmidt, 1990). Failure to correct for attenuation due to error of measurement can yield an erroneous effect size, and is critical to the credibility of the meta-analysis. Aguinis and Pierce (1998) explained that if effect sizes are not corrected for error of measurement, the meta-analyzed effect sizes have a systematic downward bias. In addition, "differential levels of measurement error across studies artificially increase the across-study variance in effect size estimates" (p. 581).

This meta-analysis corrected the mean and variance of validity coefficients across studies for attenuation due to error of measurement. If provided in the individual research study, the alpha coefficient for the reliability of the measurement instrument was coded and used in the correction of effect sizes based upon Hunter and Schmidt's (1990) method for correction of artifacts. Because alpha coefficients for measurement reliability were available in only 45 studies, the correction for error of measurement was not made on a study-by-study basis. Instead, as recommended by Hunter and Schmidt, the data was

aggregated on the measurement reliability coefficients that were available and the aggregated data was applied to all studies through a global correction method. When the average reliability was calculated in this manner, it was used to adjust any mean values calculated over all effect sizes even when each individual effect size could not be determined.

### **Range Restriction**

If effect sizes from different studies are to be compared and aggregated, then correlations due to differences in the spread (variance) on the independent variable must be controlled. Thus, correlations are directly comparable across studies only if they are computed on samples from populations with the same standard deviation on the independent variable.

According to Hunter and Schmidt (1990), the solution to range variation is to define the reference population and express all correlations in terms of that reference population. Hunter and Schmidt provided a range correction formula to compute the correlation of the sample if the standard deviation were the same as in the reference population (p. 127). Hunter and Schmidt explained that effects of range restriction are twofold: 1) range restriction produces a downward bias in study-level effect sizes, and 2) differential levels of range restriction across studies increases across-study variability in effect size estimates.

The population of interest in this meta-analysis was supervisors, managers, and leaders in organizations. The participants in the interventions in each study met the definition of “managers” as described in the criteria for inclusion of studies described in Chapter 3. This sample was believed to be representative of the population in general of

managerial leaders within organizations. Therefore, there was no reason a priori to suspect a restricted range for this meta-analysis. For this reason, studies were not corrected for range restriction as an artifact.

### **Other Potential Artifacts**

It was important to correct for as many artifacts as possible. However, there were potential artifacts that were difficult to detect, and thus were not corrected because they remain unknown. It was important to note those potential artifacts, such as error in coding raw data, errors in computing statistics or recording the numbers computed, errors in printing reliability coefficients, and errors of including extreme effect sizes. Extreme effect sizes were evaluated through careful observation of outliers.

### **Outliers**

The purpose of this meta-analysis was to arrive at a reasonable summary of quantitative findings from the research literature. This purpose would not have been served well by the inclusion of extreme effect sizes that were noticeably different from others found in the research studies. For purposes of this meta-analysis, outliers were defined as two or more standard deviations beyond the mean of its respective group, unrepresentative of the results, and were likely to have a disproportionate influence on the findings of a meta-analysis. Extreme effect sizes are typically unrepresentative of the results of the research and possibly even spurious. The Forrest plot was examined for the distribution of effect sizes to determine the presence of extreme effect sizes. No outliers were found.

### **Fixed Versus Random Effects**

There are two conceptual approaches to meta-analysis: fixed and random effects. Hunter and Schmidt (1990) indicated that if all desired levels of a variable are present in

the design, the variable is a fixed effect factor. For instance, if gender was introduced as a variable in a primary research fixed effect study, then both levels – male and female— will be present in the study. Sex would be a fixed effects factor. However, if only a sample of the levels is present in the study, then they called the variable a random effect factor (p. 405). For instance, if training content is a variable in the study, and there are a variety of topics, the content cannot be a fixed effect factor. If two or more topics are presented they might be regarded as a sample of the potential training content topics. In this case, content would be a random effect factor.

Rather than assuming that the effect size heterogeneity was due to unobserved random sources, Hunter and Schmidt's (1990) approach of partitioning effect size variance with a fixed effects model was used in this meta-analysis. This approach assumed that heterogeneity had systematic sources that were explained by variables captured in the coding process. That is, it was assumed that the excess between-study variability could be explained by the independent variables in this meta-analysis – the study and effect size descriptors.

### **Analysis of Moderating Variables**

An important advantage of meta-analysis compared to a narrative literature review strategy is that it allows for testing of the effects of moderator variables. Across-study variability in effect size estimates may be due to moderating effects as well as methodological and statistical artifacts. Once effect sizes were adjusted for sampling error and error of measurement, the distributions of effect sizes were tested for heterogeneity to determine if the effect sizes measured the same underlying phenomenon.

### **Analysis of Heterogeneity**

Hunter and Schmidt's (1990) preferred method of breaking the data into subgroups was used in this meta-analysis to determine if a study characteristic was a moderating variable. The studies were grouped by each outcome category, and a "focused test" – which attempted to explain the variation by some moderator (study characteristic) – was conducted.

This meta-analysis subscribed to Hunter and Schmidt's (1990) approach that partitions the observed effect size variability into two components: the portion attributable to subject-level sampling error and the portion attributable to other between-study differences. Hunter and Schmidt's (1990) approach of using the 75% rule was applied to this meta-analysis as the method for testing for moderator variables. The distribution was considered to be homogeneous when the sampling error accounted for 75% or more of the observed variability. The residual variance of the population was ignored unless it was at 25% or more of the observed variance. As a rule of thumb, in data where known artifacts accounted for 75% of the variance in study correlations, it was assumed that the remaining 25% was due to uncontrollable artifacts.

For computational purposes, in this meta-analysis, when the  $Q$  value for between groups ( $Q_B$ ) was more than 25% of the total  $Q$  value for the subgroup ( $Q_T$ ), the grouping variable was considered to be a moderator. This research used the analog to the analysis of variance (ANOVA) technique to group effect sizes into mutually exclusive categories on the basis of an independent variable and tested the homogeneity of effect sizes within the categories and the differences between the categories. To determine if moderators were present in this meta-analysis, an analog weighted ANOVA was run on each



subgroup within each of the three independent meta-analyses (POWC, PPWC, and SGPP). The analog to the ANOVA partitioned the total homogeneity statistic,  $Q$ , into the portion explained by the categorical variable ( $Q_B$ ) and the residual pooled within groups portion ( $Q_w$ ) (Lipsey & Wilson, 2001). This method tested the ability of the categorical variable to explain the excess effect size variability.

When testing for heterogeneity, moderators in subgroups often had only one outcome, or two outcomes from the same study. Because these small numbers of studies could potentially have given an incorrect indication of effect size, small subgroups were removed and a second ANOVA was done. This procedure determined relative effects of different treatment by study characteristics, and provided a confirmation that this disjointed subgroup failed to show as a significant subgroup, suggesting that the population measured the same phenomenon.

There was no a priori reason to believe that specific moderating variables existed in this meta-analysis. However, seven potential moderating variables were identified from the studies and tested when appropriate: the content focus of the intervention, organization type, measurement method, job classification, publication type, subjective-objective outcomes, and research design type. To determine if subjective-objective outcomes were moderators, expertise-objective and expertise-subjective outcome categories across POWC, PPWC, and SGPP designs were grouped into one meta-analysis sample. Then subjectivity-objectivity was tested as potential moderating variables.

Research design was tested as a moderator by combining studies in the expertise-objective outcome category across research design types, and running an analog of the ANOVA. In addition, expertise-subjective studies were combined across research designs

and an analog of the ANOVA was run. Research design was explored as a moderator using both expertise-objective and expertise-subjective subgroups because these were the only outcome sub-groupings with POWC, PPWC, and SGPP designs. When no moderators were detected in this meta-analysis, effect sizes were aggregated within their respective groups (defined by the study characteristic) and the meta-analysis results reported. When moderators were detected, results were reported by subgroup with no overall effect size reported. This moderator analysis process ensured construct validity of the included studies, or that the studies included in this meta-analysis were measuring the same phenomenon.

### **Summary**

Meta-analysis was used in this research to identify, aggregate, and summarize the findings of managerial leadership development studies from 1982-2001. This chapter described the method for performing this meta-analysis.

This research included formal training interventions with three research designs: POWC, PPWC, and SGPP. These research designs were used in this meta-analysis so that studies with potentially low validity, by virtue of the research design, could be considered separately from those studies potentially using more valid, experimental or quasi-experimental designs. Correlational studies were not included in the research as too few studies were located. The effect size for each individual study was based upon the statistical procedures/design used in the study.

This meta-analysis allowed for an interpretation of inconsistencies in previous results. First, the studies were corrected for statistical artifacts that may have caused the inconsistencies. The variance of the combined adjusted results was then used to

determine if substantial inconsistencies remained. If the variance was sufficiently small, there was not a need for further investigation. If the variance was not small, there were moderators that systematically influenced the results leading to inconsistencies between studies. Potential explanations of the inconsistencies were found by identifying moderator variables, partitioning of the moderator variables, and examining the variance of the subsets to determine if these variables reduced the variance, thereby explaining the inconsistent results. A moderator analysis was conducted to provide an additional segment of knowledge concerning the effectiveness of managerial leadership development program.

## **CHAPTER 4: RESULTS**

A profile of studies and the findings of this meta-analysis on the effectiveness of managerial leadership development programs from 1982-2001 are described in this chapter. The first portion of the chapter provides the results of the literature search, and a description of the studies. The description includes study characteristics such as publication year, publication type, job classification level of participants, organization type, and the country in which the intervention was performed. In addition, a description of the sample is provided by outcome categories, intervention type, content focus of the intervention, and the measurement method.

The second portion of this chapter is divided into two parts: determination of effect sizes and moderator analysis. First, findings for the three meta-analyses are presented: posttest only with control group (POWC), pretest-posttest with control group (PPWC), and single group pretest-posttest design studies (SGPP). Then, analyses of five potential moderators for each meta-analysis (POWC, PPWC, and SGPP), and two additional moderators (research design and objective-subjective outcomes) across the entire meta-analysis sample are discussed.

### **The Meta-Analysis Data Base**

An electronic search for relevant studies in ERIC, PsychInfo, and Dissertation Abstracts International databases located 6459 abstracts on the full range of managerial leadership development programs from January 1982 through December 2001. The abstracts included a large number of duplicate citations, and citations on topics, such as training methods or the developmental aspects of management jobs, that were not specifically of interest to this meta-analysis. Because the key-word combinations were

selected to be as inclusive as possible, some of the citations had nothing at all to do with managerial ranks within an organization. For example, some citations related to training of line employees or clerical support staff at the lowest levels of the organization, or cited the participants as “employees” of the organization.

The search of Web sites produced 17 articles, and the manual search of reference lists and journal indexes produced 56 additional articles of interest to this study. In addition to the computer and manual searches, contacts were made with 242 individuals, who were lead or second authors of pertinent articles, to locate unpublished studies. Through this method, institutions of higher learning and various companies were included in the literature search. One additional unpublished study was located through this process.

In total, 346 studies were located through the literature search that merited further review. Full copies of those 346 articles were obtained and read in detail to determine if they met the criteria for inclusion in this study. Two hundred fourteen (214) of the articles retrieved did not meet the study criteria as described in Chapter 3. The articles were not applicable because of one of the following reasons:

- 1) Described some theoretical aspect of management development.
- 2) Defined the training methods of an intervention.
- 3) Summarized the developmental aspects of management positions.
- 4) Described a study of naturally occurring management processes.
- 5) Defined the behavioral change of a student group who participated in an intervention.
- 6) Described an intervention of non-managerial level employees.

The authors of 26 studies were contacted for additional statistical data. One author provided the appropriate statistical information from which to determine an effect size. The remaining 25 studies were discarded from this meta-analysis.

Four articles were discarded because they were duplicates of a study published at an earlier date. The duplications that were found and the course of action were:

- 1) An author published his/her dissertation research in a scholarly journal. In this circumstance the dissertation was used as the resource for this meta-analysis, as it provided more statistical details.
- 2) An author published the same research in a refereed journal and in a conference proceeding. The study as it appeared in the refereed journal was used in this meta-analysis.
- 3) Someone other than the lead author published the study in a different format but used the same statistical analysis. For this situation, the first article published with adequate statistical data was used.
- 4) An author published the results of the same intervention at year one with a follow-up of the participants five years later. The statistical analysis from the first measurement of the intervention was used in this meta-analysis.

Thus, one hundred three (103) studies met the criteria for inclusion in this meta-analysis sample. Those 103 studies were a comprehensive group of studies with a full range of interventions that included feedback, developmental relationships, on-the-job experiences, and formal training.

As shown in Table 2, 80% of the studies located were formal training interventions (83 studies), and 13% (13 studies) were feedback interventions.

Developmental relationships that included coaching and mentoring were found in 5% (5 studies), and 2% had on-the job interventions (2 studies).

Table 2: Types of Managerial Leadership Development Interventions from 1982-2001

	Number	Frequency
Formal Training	83	80%
Feedback interventions	13	13%
Developmental relationships	5	5%
On-the-job interventions	2	2%
Total	103	100%

Feedback studies were distributed across four research design types, with three studies in three outcome subgroups in POWC design, two studies in two outcome subgroups in PPWC design, six studies in three outcome subgroups in SGPP design, and two studies in two outcome subgroups with correlation design (CORR). Because feedback interventions are relatively new in the literature and there is not enough information regarding their effectiveness, there were only a small number of studies available for this meta-analysis. For these reasons, feedback interventions were discarded from this meta-analytic research. In addition, there were too few studies with developmental relationships and on-the-job experiences to adequately perform a meaningful meta-analysis. Therefore, this meta-analysis was reduced to 83 studies with formal training interventions as the final meta-analysis sample. A reference list and a summary of the 83 studies in the meta-analysis sample are provided in Appendix B.

## Profile of Managerial Leadership Development Studies

### Publication Characteristics

Table 3 summarizes the publication sources from which studies in this meta-analytic research were obtained. The studies, published in 27 different professional

Table 3: Publication Sources of Managerial Leadership Development Studies from 1982-2002

Publication	# Studies	Frequency	# Effect Sizes	Frequency
Academy of Management Journal	1	1%	6	4%
Employee Assistance Quarterly	2	3%	2	1%
Evaluation Review	1	1%	1	1%
Group & Organization Studies	1	1%	2	1%
Human Resource Development Quarterly	5	7%	7	5%
Journal of Applied Behavioral Science	1	1%	1	1%
Journal of Business and Psychology	1	1%	2	1%
Journal of Classroom Instruction	1	1%	1	1%
Journal of Community Psychology	1	1%	1	1%
Journal of Continuing Education Nursing	1	1%	1	1%
Journal of Educational Research	1	1%	1	1%
Journal of Employee Assistance Research	1	1%	3	2%
Journal of Managerial Psychology	1	1%	1	1%
Journal of Occupational Psychology	1	1%	2	1%
Journal of Organizational Behavior	2	3%	2	1%
Management				
Journal of Parks and Recreation	1	1%	2	1%
Administration				
Journal of Organizational Behavior	1	1%	2	1%
Journal of Applied Psychology	8	10%	19	13%
Leadership Quarterly	1	1%	6	4%
Management Education and Development	1	1%	2	1%
Organizational Behavior & Human				
Decision Processes	1	1%	2	1%
Organizational Behavior & Human	1	1%	3	2%
Performance				
Personnel Psychology	7	9%	15	10%
Psychology Reports	1	1%	2	2%
Public Productivity & Management Review	1	1%	1	1%
South African Journal of Psychology	1	1%	2	2%
Training and Development Journal	2	3%	3	2%
Dissertations	32	39%	51	34%
Other unpublished studies	4	5%	4	3%
Total	83	100%	150	100%



journals, produced 150 effect sizes. The *Journal of Applied Psychology* was the most common source, contributing eight studies with 19 effect sizes. *Personnel Psychology* contributed seven studies with 15 effect sizes, and *Human Resource Development Quarterly* contributed five studies with seven effect sizes. Thirty-two (32) studies were doctoral dissertations, two were from proceedings of professional meetings and two were unpublished manuscripts. Dissertations contributed 51 effect sizes (34%) to this meta-analysis.

### **Job Classification Level**

The primary focus of managerial leadership development interventions was on the supervisor/foreman job classification level, considered to be entry-level management for purposes of this meta-analysis (See Table 4). Twenty-seven (27) studies focusing on entry-level management generated 35% of the effect sizes. Thirty-two (32) studies focused on a mixed management group of participants and generated 35% of the effect sizes. Top management and mid-management level interventions were each found in 13 and 11 studies respectively. For this research, studies with school principals were considered to be top management level interventions.

Table 4: Frequency of Managerial Leadership Development Interventions by Job Classification Level of Participants

Level	# Studies	Frequency	# Effect Sizes	Frequency
Supervisor/Foreman	27	33%	52	35%
Mid Manager	11	13%	21	14%
Top Management	13	16%	25	16%
Mixed management	32	38%	52	35%
Total	83	100%	150	100%

## **Organization Type**

Table 5 provides a break down of interventions by organization type, or the setting in which the intervention occurred. Eighteen (18) studies described interventions in an education setting and contributed 27 effect sizes to this meta-analysis. Interventions in business and industry were captured in 16 studies that contributed 34 effect sizes. Thirteen (13) studies (24 effect sizes) occurred in government settings, and nine studies (13 effect sizes) came from studies with a medical orientation. Organization types in the business and industry subcategory were automotive, financial, manufacturing, technology and utilities. Twenty-three (23) studies failed to describe the organization type of the intervention.

Table 5: Frequency of Managerial Leadership Development Programs by Organization Type

Organization Type	# Studies	Frequency	# Effect Sizes	Frequency
Business/Industry	16	19%	34	24%
<i>Automotive</i>	3	4%	7	5%
<i>Financial</i>	2	2%	4	3%
<i>Manufacturing</i>	4	5%	11	7%
<i>Technology</i>	5	6%	7	5%
<i>Utilities</i>	2	2%	5	4%
Education	18	21%	27	18%
Government	13	15%	24	16%
Medical	9	11%	13	8%
Military	6	7%	14	9%
Other/unknown	23	27%	38	25%
Total	85	100%	150	100%

Note. One study analyzed interventions from three different industries.

### **Country In Which Intervention Occurred**

Table 6 provides a summary of the countries in which managerial leadership development programs in this research occurred. Fifty-six (56) studies described interventions that occurred in the United States, with those studies providing 101 (67%) of the effect sizes. Twenty (20) studies contributing 34 (23%) effect sizes occurred outside the United States. If the study did not specifically identify the location, but indicated that the company was multinational, the intervention was considered to have occurred in both the United States and countries outside the United States. Three studies described interventions in multinational companies.

Table 6: Frequency of Managerial Leadership Development Interventions by Country

Country	# Studies	Frequency	# Effect Sizes	Frequency
Non-US	20	24%	34	23%
US	56	67%	101	67%
Both US/Non-US	3	4%	9	6%
Unknown	4	5%	6	4%
Total	83	100%	150	100%

### **Content Focus of Formal Training Interventions**

The interventions in this research were grouped into six training content areas (See Table 7). Human relations content was the most prevalent content focus with 44 studies that contributed 80 (53%) of the effect sizes. Twenty-nine (29) studies with 48 effect sizes had a general management training content focus. Other content focus areas were employee performance, job and work design, problem solving/decision making, and strategic stewardship.

Table 7: Frequency of Managerial Leadership Development Programs by Content Category

Content Category	# Studies	Frequency	# Effect Sizes	Frequency
Employee Performance	2	2%	5	4%
General Management	29	35%	48	32%
Human Relations	44	53%	80	53%
Job & Work Redesign	2	2%	5	3%
Problem Solving/Decision Making	3	4%	5	3%
Strategic Stewardship	3	4%	5	4%
Total	83	100%	150	100%

### **Outcome Levels of Formal Training Interventions**

Eighty-three (83) studies in this meta-analysis sample contributed a total of 150 effect sizes, an average of 1.8 effect sizes per study. Table 8 shows that 52 studies (40%) had expertise-subjective outcomes and contributed 63 effect sizes. Forty-five studies (34%) with expertise-objective outcomes produced 52 effect sizes. It is interesting to note that performance-level (system or financial) studies were sparse in this meta-analysis. Eleven (11) studies measured system objectives, and only one study had financial outcomes. Nine (9) studies had simultaneous independent training groups. In those studies, the outcomes from each independent training group were entered separately into the meta-analysis.

Learning has always been an important outcome of training. Training professionals typically develop objective tests for cognitive knowledge to measure trainees' learning objectively. Twenty-two studies (17%) in this meta-analysis measured knowledge outcomes and contributed 24 effect sizes. On the other hand, 97 studies had interventions with expertise outcomes, producing 115 (75%) of the effect sizes.

Table 8: Frequency of Managerial Leadership Development Studies by Outcome Category

Outcome Category	# Studies	Frequency	# Effect Sizes	Frequency
Knowledge-Objective	17	13%	19	13%
Knowledge-Subjective	5	4%	5	3%
Expertise-Objective	45	34%	52	34%
Expertise-Subjective	52	40%	63	41%
System-Objective	9	7%	9	6%
System-Subjective	2	1%	2	1%
Financial-Objective	1	1%	1	1%
Total	131	100%	150	100%

Note. Many studies have multiple outcomes, in multiple outcome subgroups.

### **Measurement Methods**

Table 9 provides a summary of studies by measurement method. Self-assessments, which were considered subjective measurements, were used in 55 studies (49.5%) in this meta-analysis. Other measurements including objective tests, other person assessments, and company records occurred in 56 studies (50.5%). Seventy (70) effect sizes were produced from the self-assessment studies and 80 from studies with other types of measurements. Twenty-eight (28) studies used a combination of self and other measurement instruments to measure the effectiveness of managerial leadership development. It was noted that 46 different measurement instruments were used in

Table 9: Frequency of Managerial Leadership Development Studies by Measurement Method

Method	# Studies	Frequency	# Effect Sizes	Frequency
Self	55	49.5%	70	47%
Other	56	50.5%	80	53%
Total	111	100%	150	100%

Note. Twenty-eight (28) studies have a combination of self and other measurements.

studies in this meta-analysis. Cronbach's alpha coefficients were provided for 45 of the studies with an average reliability over those studies of .893.

### **Implication of Research Design in Meta-Analysis**

Although this study was initially intended to be a comprehensive meta-analysis on the overall effectiveness of *all* managerial leadership development programs from 1982-2001, it actually incorporated only formal training interventions. The meta-analysis sample was separated by four research designs, based upon the type of research design used in individual studies. The four research designs as shown in Table 10 are pretest only with control group (POWC), pretest-posttest with control group (PPWC), single group pretest-posttest (SGPP), and correlational (CORR).

Each of the four research designs had the potential of having study outcomes in eight categories: knowledge-objective and subjective; expertise-objective and subjective; system-objective and subjective; and financial-objective and subjective. Thus, across the four types of research designs, a total of 32 possible outcome subgroups with effect sizes were initially considered (four designs X eight outcome subgroups). Table 10 indicates the number of effect sizes coded per outcome subgroup for each type of research design.

Table 10: Initial Number of Managerial Leadership Development Studies and Effect Sizes (k) per Outcome Subgroup

	POWC		PPWC		SGPP		CORR	
	Studies	k	Studies	k	Studies	k	Studies	k
Knowledge-Objective	7	9	4	4	6	6	1	1
Knowledge-Subjective	1	1	0	0	0	0	3	3
Expertise-Objective	12	15	19	23	14	14	1	1
Expertise-Subjective	19	24	18	19	13	15	2	3
System Objective	5	5	2	2	2	2	0	0
System-Subjective	1	1	1	1	0	0	0	0
Financial-Objective	0	0	1	1	0	0	0	0
Financial-Subjective	0	0	0	0	0	0	0	0
Total	32	55	34	50	27	37	6	8

While technically a meta-analysis can be conducted with as few as two effect sizes (Hunter & Schmidt, 1990), the decision was made to conduct this meta-analysis with comparisons of six or more effect sizes, as that was a natural separation point in the study. There has been an extensive review by leading meta-analytic researchers examining the statistical power of meta-analysis (Sackett, in press). The primary caution is to not draw strong conclusions from a small meta-analysis sample although there are no definitive rules for defining “small.” Sackett explained that power depends on the conditions simulated, the Type I error rate, and the magnitude of validity difference one is interested in detecting; thus, there is no single value that can be used to represent power. Therefore, the decision was made to not analyze the smaller outcome subgroups.

Several outcomes subgroups in this meta-analysis had fewer than six effect sizes. In those situations the following measures were taken:

- 1) Two cells (outcome subgroups) with less than six effect sizes in the PPWC design data set were moved to the same outcome subgroup in the POWC design data set by discarding the pretest-posttest data and using only the control group comparison data. This was done in order to include studies that would otherwise not have been included in this meta-analysis. Those situations are as follows:
  - PPWC knowledge-objective outcomes were merged with POWC knowledge-objective outcomes.
  - PPWC system-objective outcomes were merged with POWC system-objective outcomes.
- 2) Effect sizes in subgroups with less than six effect sizes that could not be logically combined were not used in the aggregation of studies in this meta-analytic

research. Therefore, the following categories were disregarded because too few studies were located:

- POWC knowledge-subjective
- POWC system-subjective
- PPWC system-subjective
- PPWC financial-objective
- SGPP system-objective

3) Correlational studies (CORR) were dropped because of the small number of studies located. Only six studies (eight effect sizes) existed in the entire correlational data set. Three studies were located with knowledge-subjective outcomes, one with knowledge-objective, two with expertise-subjective, and one with expertise-objective outcomes.

Table 11 shows the number of studies and effect sizes per outcome subgroup after small cell sizes were adjusted. The categories are by dependent variable separated by three study designs: POWC, PPWC, and SGPP. The four outcome categories included for

Table 11: Final Meta-Analysis Sample of Managerial Leadership Development Studies from 1982-2001 by Outcome Subgroups

	POWC		PPWC		SGPP		CORR	
	Studies	k	Studies	k	Studies	k	Studies	k
Knowledge-Objective	11	13	0	0	6	6	0	0
Knowledge-Subjective	0	0	0	0	0	0	0	
Expertise-Objective	12	15	19	23	14	14	0	0
Expertise-Subjective	19	24	18	19	13	15	0	0
System Objective	7	7	0	0	0	0	0	0
System-Subjective	0	0	0	0	0	0	0	0
Financial-Objective	0	0	0	0	0	0	0	0
Financial-Subjective	0	0	0	0	0	0	0	0
Total	36	59	26	42	25	35	0	0

Note. k = effect size.



POWC design studies were knowledge-objective, expertise-objective, expertise-subjective, and system-objective. Two outcome categories for PPWC studies were included in this meta-analysis: expertise-subjective and expertise-objective. For studies with SGPP designs, knowledge-objective, expertise-subjective, and expertise-objective outcomes (or three total outcome categories) were included in this meta-analysis. Thus, the final sample of studies analyzed was 83.

Table 12 presents a summary of the number of studies and effect sizes per outcome category in the final set of studies included in this meta-analysis. A total of 136 effect sizes were aggregated in this meta-analysis. Expertise-subjective outcomes comprised the largest category with 50 studies producing 58 effect sizes (43%). The next largest was expertise-objective with 45 studies and 52 effect sizes. Knowledge-objective and system-objective outcome categories were considerably smaller, with 17 and five studies respectively. Findings for outcome subgroups for each type of research design (i.e., POWC, PPWC, and SGPP) are discussed in the following sections.

Table 12: Frequency of Managerial Leadership Development Studies and Effect Sizes per Outcome Subgroup in Final Meta-Analysis Sample

Outcome Subgroup	# Studies	Frequency	# Effect Sizes	Frequency
Knowledge-Objective	17	14%	19	14%
Expertise-Objective	45	38%	52	38%
Expertise-Subjective	50	42%	58	43%
System-Objective	7	6%	7	5%
Total	119	100%	136	100%

Note. Many studies have multiple outcomes, in multiple outcome subgroups.

### **Posttest Only With Control Group (POWC) Meta-Analysis**

The POWC design studies compare posttest scores for the trained and untrained groups. The POWC design was unique in this research, as it was the only one of the three independent meta-analyses that did not contain pretest scores. Essentially, an effect size in POWC design studies was the normalized difference between trained and untrained group outcomes (Carlson & Schmidt, 1999).

Thirty-six (36) studies were incorporated in this POWC meta-analysis of the effectiveness of managerial leadership development. Those studies generated 59 effect sizes with a total of 3,335 subjects (See Table 13). It is important to note that the expertise-subjective outcome subgroup was the largest number with 1,335 subjects. Thirteen (13) studies had more than one category of dependent variable and five studies had more than one experimental treatment.

Four outcome subgroups were analyzed in the POWC meta-analysis: expertise-objective, expertise-subjective, knowledge-objective, and system-objective. Twelve (12) studies contributed 15 effect sizes in the expertise-objective subgroup, 19 studies contributed 24 effect sizes in the expertise-subjective outcome subgroup, 11 studies with knowledge-objective outcomes produced 13 effect sizes, and seven studies produced seven system-objective outcomes. It should be noted that the effect sizes in the POWC data set vary greatly (from  $-1.39$  to  $2.02$ ).

It is important to note that, as described in the previous section, four of the effect sizes in the knowledge-objective outcomes subgroup and two effect sizes in the system-objective outcomes subgroup were originally a part of the PPWC meta-analysis, but were

Table 13: Managerial Leadership Development Studies with POWC Research Design

Citation (Year)	N		Effect Size	Std Err	95% Confidence Interval		Org Type	Job Class	Content
					Lower Limit	Upper Limit			
	Training	Control							
EXPERTISE-OBJECTIVE									
Alsamani (1997)	38	31	1.39	.27	.85	1.93	G	M	G
Bendo (1984)	63	63	.39	.18	.03	.74	O	E	H
Clark et al (1985)	19	8	.84	.44	-.06	1.74	Me	O	H
Davis/Mount (1984)	66	96	-1.39	.18	-1.75	-1.05	O	M	E
Davis/Mount (1984)	104	98	.40	.14	.12	.68	O	M	E
DePiano/McClure (1987)	23	60	.74	.25	.23	1.24	E	T	H
Dvir et al (2001)	23	17	-.90	.34	-1.56	-.22	Mi	E	S
Earley (1987)	20	20	2.01	.39	1.22	2.79	Ma	E	G
Earley (1987)	20	20	1.03	.34	.34	1.71	Ma	E	G
Earley (1987)	20	20	1.44	.35	.72	2.16	Ma	E	G
Eden (1986)	7	9	.89	.53	-.24	2.02	Mi	O	J
Henry (1983)	156	156	.24	.11	.01	.46	T	E	H
Josefowitz (1984)	65	31	-.08	.22	-.52	.35	T	E	G
Scandura/Graen (1984)	21	57	.51	.26	-.01	1.02	G	E	H
Thoms/Klein (1994)	60	60	.41	.29	-.11	.96	Me	O	H
SUBTOTAL (15)	705	686	.33	.08	.18	.49			
EXPERTISE-SUBJECTIVE									
Alsamani (1997)	38	31	1.00	.26	.49	1.51	G	M	G
Briddell (1986)	24	24	.02	.29	-.56	.60	E	O	G
Colan/Schneider (1992)	184	50	.26	.16	-.05	.58	U	E	H
Dvir et al (2001)	27	18	.52	.31	-.11	1.14	Mi	E	S
Earley (1987)	20	20	1.23	.34	.53	1.93	Ma	E	G
Earley (1987)	20	20	1.15	.34	.46	1.84	Ma	E	G
Earley (1987)	20	20	1.53	.36	.80	2.26	Ma	E	G
Eden (1986)	7	9	.89	.53	-.24	2.02	Mi	O	J
Eden et al (2000)	369	599	.13	.07	-.001	.26	Mi	O	H
Fuller (1985)	24	24	-.48	.29	-1.07	.11	E	M	H
Gerstein et al (1989)	112	112	.23	.13	-.03	.50	A	O	G
Harrison (1992)	11	12	.62	.43	-.27	1.50	Mi	O	H
Harrison (1992)	11	12	.02	.42	-.85	.88	Mi	O	H
Henry (1983)	156	156	.28	.11	.06	.51	T	E	H
Ivancevich (1992)	15	15	.62	.37	-.14	1.39	O	O	H
Ivancevich (1992)	15	15	.74	.38	-.04	1.51	O	O	H
Ivancevich (1992)	15	15	.43	.37	-.33	1.18	O	O	H
Maurer/Fay (1988)	21	21	-.44	.31	-1.07	.19	Me	O	H
Moxnes/Eilerten (1991)	77	133	.14	.14	-.14	.42	U	E	G
Reaves (1993)	25	20	1.01	.32	.37	1.65	E	O	H
Scandura/Graen (1984)	21	57	.38	.26	-.13	.89	G	E	H
Tziner et al (1991)	45	36	.56	.23	.10	1.01	Mi	E	H
Williams (1992)	49	30	.42	.23	-.32	.60	G	E	G
Young /Dixon (1995)	29	40	-.78	.25	.27	1.28	O	T	G
SUBTOTAL (24)	1,335	1,489	.30	.05	.12	.42			
KNOWLEDGE-OBJECTIVE									
Davis/Mount (1984)	88	122	1.60	.16	1.28	1.91	O	M	E
Davis/Mount (1984)	135	122	1.19	.14	.92	1.45	O	M	E
DeNisi/Peters (1996)	66	22	.65	.25	.15	1.15	Ma	E	H
Haccoun/Hamtiaux (1994)	42	24	.83	.27	.29	1.36	E	M	H
Harrison (1992)	11	12	1.44	.47	.46	2.41	Mi	O	H
Harrison (1992)	11	12	1.56	.48	.57	2.55	Mi	O	H
Maurer/Fay (1988)	212	21	1.10	.23	.63	1.56	Me	O	H
May/Kahnweiler (2000)	19	19	1.01	.34	.31	1.71	Ma	E	H

(Table cont.)

		95% Confidence Interval							
Citation (Year)	N		Effect Size	Std Err	Lower Limit	Upper Limit	Org Type	Job Class	Content
Russell et al (1984)	19	11	1.39	.42	.53	2.25	A	M	H
Thoms/Klein (1994)	64	64	.32	.27	-.18	.85	Me	O	H
Tziner et al (1991)	45	36	.72	.23	.26	1.17	Mi	E	H
Wolf (1996)	144	144	.76	.12	.52	1.00	Me	O	H
Yaworsky (1994)	21	21	.61	.32	-.03	1.25	G	E	G
SUBTOTAL (13)	877	630	.96	.07	.82	1.12			
SYSTEM-OBJECTIVE									
Bankston (1993)	13	13	.79	.31	.42	.93	E	T	H
Colan/Schneider (1992)	184	50	.30	.16	-.01	.62	U	E	H
Graen et al (1982)	36	95	.60	.20	.20	.99	G	O	J
Hill (1992)	25	27	.10	.27	-.46	.66	E	T	H
Posner (1982)	34	34	.02	.24	-.46	.50	G	E	H
Scandura/Graen (1984)	21	57	.67	.26	.15	1.19	G	E	H
Urban et al (1985)	105	105	.43	.14	.16	.71	O	E	G
SUBTOTAL (7)	418	381	.39	.10	.19	.59			
COMBINED (59)	3,335	3,186							

**Note.** Organization Type (Org Type): G = Government; Ma = Manufacturing; T = Technology; E = Education; Mi = Military; U = Utilities; Me = Medical; A = Automotive; and O = Other Unknown. Job Classification (Job Class): T = Top Management; M = Mid-Manager; E = Entry Level; and O= Mixed Groups. Training Content (Content): G = General Management; H = Human Relations; S = Strategic Stewardship; E = Employee Performance; and J = Job & Work Redesign

combined with the POWC meta-analysis. Therefore, this POWC meta-analysis incorporated six additional effect sizes that would not have originally been included.

### **POWC Effect Sizes**

Carlson and Schmidt's formulas (1999, p. 852) were used to calculate the magnitude of the effect of the managerial leadership development programs in POWC studies from the statistics available in the individual research studies. The study effect sizes were adjusted for sampling error variance and error of measurement as described in Chapter 3. Table 14 provides effect sizes for the POWC data set grouped by the four outcome subgroups (i.e., expertise-objective, expertise-subjective, knowledge-objective, and system-objective). As indicated in Table 14, the overall effect size of expertise-objective studies was .33, based upon 15 effect sizes and 705 subjects. Expertise-

Table 14: Effect Sizes of Managerial Leadership Development Interventions per Outcome Subgroup for POWC Design Studies

Outcome Subgroup	k	N	Effect Size	Standard Error	95% Confidence Interval	
					Lower Limit	Upper Limit
Expertise-Objective	15	705	.33	.08	.18	.49
Expertise-Subjective	24	1,335	.30	.05	.12	.42
Knowledge-Objective	13	877	.96	.07	.82	1.12
System-Objective	7	418	.39	.10	.19	.59
Total	59	3,335				

subjective studies (24 effect sizes) produced an average effect size of .30 from 1,335 subjects. The knowledge-objective overall effect size was .96, aggregated from 13 effect sizes and 877 subjects. System-objective studies produced an overall effect size of .39 from seven effect sizes and 418 subjects.

Results of an analogue weighted ANOVA of the POWC data set indicated that variance in the data set was 27.4%, an amount greater than that attributed to sampling error. Further explanation regarding possible moderators in POWC studies is provided later in this chapter.

### **Pretest-Posttest with Control Group (PPWC) Meta-Analysis**

The inclusion of a control group and the use of the pre-training dependent variable standard deviation are believed by researchers to make the PPWC design superior (Carlson & Schmidt, 1999). In addition, Carlson and Schmidt believe that effect sizes from PPWC designs are based upon more information about training, and should more closely approximate the true effect size.

Table 15: Managerial Leadership Development Studies with PPWC Research Design

Citation (Year)	N		Effect Size	Std Err	95% Confidence Interval		Org Type	Job Class	Content
					Lower Limit	Upper Limit			
	Training	Control							
<b>EXPERTISE-OBJECTIVE</b>									
Bankston (1993)	13	15	.57	.38	-.23	1.36	E	T	H
Birkenbach et al (1984)	25	25	.89	.30	.29	1.48	Ma	E	H
Deci et al (1989)	235	177	.49	.10	.29	.69	O	O	H
Devlin-Scherer et al (1997)	162	162	.04	.11	-.18	.25	E	O	H
Eden et al (2000)	17	17	1.29	.38	.52	2.05	E	T	H
Eden et al (2000)	398	583	.26	.07	.13	.39	Mi	O	H
Eden et al (2000)	12	10	.27	.43	-.62	1.17	F	T	H
Frost (1996)	33	17	.37	.30	-.24	.97	G	O	H
Frost (1996)	28	17	-.45	.31	-1.07	.18	G	O	H
Graen et al (1982)	37	95	.74	.20	.35	1.14	G	O	J
Graen et al (1982)	37	95	.51	.20	.11	.89	G	O	J
Mattox (1985)	18	18	-.05	.33	-.73	.63	O	O	G
May/Kahnweiler (2000)	19	19	1.01	.35	.31	1.71	Ma	E	H
Nelson (1990)	30	14	.04	.32	-.61	.69	E	T	G
Niska (1991)	13	13	1.12	.42	.25	2.00	E	T	G
Rosti/Shipper (1998)	27	26	.41	.28	-.15	.97	O	M	G
Russell et al (1984)	9	16	.23	-	-.64	1.09	A	M	H
Savan (1983)	25	25	.07	.28	-.50	.64	O	E	H
Smith et al (1992)	14	14	1.22	.41	.38	2.07	E	O	H
Sniderman (1992)	59	146	.20	.20	-.19	.59	O	O	H
Steele (1984)	220	219	.03	.10	-.16	.22	T	M	H
Tharenou/Lyndon (1990)	50	50	.78	.21	.37	1.19	G	E	G
Yaworsky (1994)	21	21	.09	.31	-.54	.71	G	E	G
SUBTOTAL (23)	1,502	1,794	.32	.05	.22	.43			
<b>EXPERTISE-SUBJECTIVE</b>									
Bankston (1993)	13	15	.33	.33	-.45	1.11	E	T	H
Barling et al (1996)	9	11	.42	.45	-.53	1.38	F	T	S
Birkenbach et al (1984)	25	25	.78	.29	.19	1.37	Ma	E	H
Cato (1990)	40	40	.29	.22	-.16	.74	G	O	G
Clark (1990)	31	63	.11	.22	-.33	.54	E	E	H
Deci et al (1989)	8	13	.20	.45	-.74	1.15	O	O	H
Eden et al (2000)	17	17	.60	.35	-.11	1.31	E	T	H
Eden et al (2000)	21	23	.57	.31	-.05	1.20	F	T	H
Edwards (1992)	29	39	1.22	.27	.69	1.76	E	O	P
Hill (1992)	25	27	-.13	.28	-.69	.42	E	T	H
Lawrence/Wiswll (1993)	33	32	.29	.25	-.20	.79	G	O	G
Mattox (1985)	18	18	-.15	.33	-.83	.53	O	O	G
Nelson (1990)	30	14	.17	.32	-.48	.82	E	T	G
Niska (1991)	13	13	1.67	.46	.73	2.62	E	T	G
Russell et al (1984)	11	17	-.22	.38	-1.01	.58	A	M	H
Savan (1983)	25	25	-.01	.28	-.57	.56	O	E	H
Steele (1984)	50	50	.09	.20	-.30	.49	T	M	H
Tharenou/Lyndon (1990)	50	50	1.06	.21	.64	1.49	G	E	G
Yaworsky (1994)	21	21	.09	.31	-.54	.71	G	E	G
SUBTOTAL (19)	469	417	.40	.10	.20	.61			
COMBINED (42)	1,971	2,190							

Note. Organization Type (Org Type): G = Government; T = Technology; E = Education; Mi = Military; F = Financial; A = Automotive; and O = Other Unknown. Job Classification (Job Class): T = Top Management; M = Mid-Manager; E = Entry Level; and O = Mixed Groups. Training Content (Content): G = General Management; H = Human Relations; S = Strategic Stewardship; J = Job & Work Redesign; and P = Problem Solving.

Twenty-four (26) studies were incorporated in the PPWC meta-analysis. Those studies generated 42 effect sizes with a total of 2,190 subjects. The meta-analysis of PPWC studies aggregated expertise-objective and expertise-subjective outcomes only, as those were the only two categories with six or more effect sizes.

Nineteen (19) studies contributed 23 effect sizes with expertise-objective outcomes, and 18 studies contributed 19 effect sizes with expertise-subjective outcomes. Table 15 shows that 1,502 subjects were included in the expertise-objective outcome subgroup and 469 in expertise-subjective. Eleven (11) studies had more than one type of dependent variable measured and two studies had more than one experimental treatment. It should be noted that the effect sizes in PPWC studies vary greatly (from -.45 to 1.67).

### **PPWC Effect Sizes**

Given the statistics available from the research studies, Carlson and Schmidt's (1999, p. 852) formulas were used to estimate the magnitude of the effect of the managerial leadership development programs in PPWC studies. The study effect sizes were adjusted for sampling error variance and error of measurement as described in Chapter 3.

The average effect sizes found for PPWC studies are reported in Table 16. The overall effect size for expertise-objective studies was .32 (aggregated from 23 effect sizes and 1,502 subjects). The observed effect size for expertise-subjective studies was .40, determined from 19 effect sizes and 469 subjects. It should be noted that the number of subjects is small considering the expansiveness of this meta-analysis.

The data set was grouped by the two outcome categories (i.e., expertise-objective and expertise-subjective) and an analogue weighted ANOVA was run for the PPWC data

Table 16: Effect Sizes of Managerial Leadership Development Interventions per Outcome Subgroup for PPWC Design Studies

Outcome Subgroup	k	N	Effect Size	Standard Error	95% Confidence Interval	
					Lower Limit	Upper Limit
Expertise-Objective	23	1,502	.32	.05	.22	.43
Expertise-Subjective	19	469	.40	.10	.20	.61
Total	42	1,971				

set by outcome category. Variance within the data set appeared to be within the normal expectations of variance as a result of random error. Therefore, potential moderators did not appear to exist in this PPWC meta-analysis research. However, further research is advised because of the low power of this analysis to detect moderators. Although not found in this research, undetermined potential moderators could possibly exist and must be determined through future research. Further explanation regarding potential moderators in PPWC studies is provided later in this chapter.

### **Single Group Pretest-Posttest (SGPP) Meta-Analysis**

The SGPP research design is often the only design possible to evaluate training programs (Hunter & Schmidt, 1990) and to examine behavior change in the participants. However, SGPP design is normally not included in meta-analyses because it is believed to create upward bias of the mean treatment effect due to threats to internal and external validity. But, in this meta-analysis the pre-training standard deviation of the mean treatment effect was used in computing the effect size and was believed to reduce any bias (Carlson & Schmidt, 1999). In addition, Hunter and Schmidt recommend including this design, as it allows for easier detection of treatment by subject interaction, or individual participant differences related to the intervention.



A total of 25 studies were incorporated in this SGPP meta-analysis of the effectiveness of managerial leadership development programs. These studies generated 35 effect sizes with a total of 4,284 subjects on which to base the aggregation. It should be noted that the effect sizes in the SGPP data set vary greatly from (-.28 to 2.10). Table 17 shows that fourteen (14) studies were included with expertise-objective outcomes producing 14 effect sizes. Thirteen (13) studies contributed 15 effect sizes in the expertise-subjective outcome subgroup, and six studies produced six effect sizes in the knowledge-outcomes subgroup of the SGPP meta-analysis. Twelve (12) studies had more than one type of dependent variable and two studies had more than one experimental treatment. It should be noted that two studies initially existed in the system-objective outcome subgroup, but were discarded because not enough studies were available to adequately perform an analysis as stipulated by the standards of this meta-analysis (See Tables 10 and 11). The final aggregation of SGPP studies was based on 25 studies that generated 35 effect sizes.

### **SGPP Effect Sizes**

Given the statistics available from the research studies, Carlson and Schmidt's (1999, p. 852) formulas were used to estimate the magnitude of the effect of the managerial leadership development programs in SGPP studies. The study effect sizes were adjusted for sampling error variance and error of measurement as described in Chapter 3.

Average effect sizes for SGPP outcome subgroups are shown in Table 18. The observed effect size for expertise-objective outcomes was 1.01, averaged across 14 effect

Table 17: Managerial Leadership Development Studies with SGPP Research Design

95% Confidence Interval								
Citation (Year)	N	Effect Size	Std Err	Lower Limit	Upper Limit	Org Type	Job Class	Content
Training								
EXPERTISE-OBJECTIVE								
Bruwelheide/Duncan (1996)	12	1.43	.46	.48	2.38	O	O	H
Donohoe et al (1997)	89	1.66	.17	1.32	2.00	G	O	H
Dorfman et al (1986)	121	.09	.13	-.16	.34	E	E	E
Jalbert et al (2000)	39	-.28	.23	-.73	.18	T	T	G
Katzenmeyer (1988)	50	.41	.20	.01	.81	E	T	H
Larsen (1983)	12	.85	.40	.03	1.67	Me	M	G
McCauley/Hughes-James (1994)	38	.25	.23	-.21	.71	E	T	G
Paquet et al (1987)	22	.60	.31	-.02	1.22	O	O	G
Shipper/Neck (1990)	10	.78	.46	-.19	1.76	Me	O	H
Tesoro (1991)	11	.44	.43	-.46	1.34	O	O	H
Tracey et al (1995)	104	1.25	.15	.95	1.55	O	E	G
Warr/Bunce (1995)	106	.16	.14	-.11	.43	O	E	G
Woods (1987)	40	.06	.22	-.38	.51	E	O	H
Yang (1988)	350	1.39	.08	1.23	1.56	G	O	G
SUBTOTAL (14)	1,004	1.01	.06	.87	1.15			
EXPERTISE-SUBJECTIVE								
Faerman/Ban (1993)	1,363	.28	.04	.21	.36	G	E	H
Innami (1994)	112	1.71	.16	1.40	2.02	Me	E	P
Innami (1994)	112	.29	.13	.02	.55	Me	E	P
Katzenmeyer (1988)	50	.69	.21	.28	1.10	E	T	H
Lafferty (1998)	233	.34	.09	.16	.53	Mi	M	G
Lafferty (1998)	282	.29	.08	.12	.45	Mi	M	G
Larkin (1996)	23	.87	.31	.25	1.49	Me	O	H
Martineau (1995)	52	.39	.20	-.01	.78	O	E	H
McCauley/Hughes-James (1994)	38	.61	.23	.14	1.08	E	T	G
Robertson (1992)	160	.04	.11	-.18	.26	O	O	G
Sogunro (1997)	29	2.10	.33	1.44	2.75	O	O	G
Tenorio (1996)	19	.84	.34	.15	1.52	T	E	G
Thoms/Greenberger (1998)	105	.52	.17	.17	.86	O	O	S
Werle (1985)	20	.56	.32	-.09	1.22	O	M	G
Woods (1987)	40	.34	.23	-.11	.79	E	O	H
SUBTOTAL (15)	2,638	.38	.04	.30	.46			
KNOWLEDGE-OBJECTIVE								
Couture (1987)	13	1.06	.42	.19	1.92	Me	E	H
Larsen (1983)	9	1.22	.51	.13	2.31	Me	M	G
Martineau (1995)	67	.66	.18	.31	1.01	O	E	H
Tesoro (1991)	99	1.59	.16	1.27	1.91	O	O	H
Tracey et al (1995)	104	1.66	.16	1.34	1.97	O	E	G
Yang (1988)	350	1.54	.09	1.37	1.71	G	O	G
SUBTOTAL (6)	642	1.36	.08	1.18	1.56			
COMBINED (35)	4,284							

Note. Organization Type (Org Type): G = Government; T = Technology; E = Education; Mi = Military; Me = Medical; and O = Other Unknown. Job Classification (Job Class): T = Top Management; M = Mid=Manager; E = Entry Level; and O = Mixed Groups. Training Content (Content): H = Human Relations; E = Employee Performance; P = Problem Solving; G = General Management; and S = Strategic Stewardship.

sizes and 1,004 subjects. The effect size for expertise-subjective outcomes was .38, aggregated over 15 effect sizes and 2,638 subjects. The magnitude of knowledge-objective outcomes was 1.37 from six effect sizes and 642 subjects. While the knowledge-objective outcomes appear to be highly effective, it should be pointed out that the results were obtained from a very small sample.

Table 18: Effect Sizes of Managerial Leadership Development Interventions per Outcome Subgroup for SGPP Design Studies

Source	k	N	Effect Size	Standard Error	95% Confidence Interval	
					Lower Limit	Upper Limit
Expertise-Objective	14	1,004	1.01	.06	.87	1.15
Expertise-Subjective	15	2,638	.38	.04	.30	.46
Knowledge-Objective	6	642	1.37	.08	1.18	1.56
Total	35	4,284				

The SGPP data set was grouped by the three outcome categories found in the studies: knowledge-objective, expertise-objective, and expertise-subjective. An analogue weighted ANOVA for the SGPP data set by outcome category. Moderators appeared to be prevalent in the SGPP data set because, according to Hunter and Schmidt's (1990) 75% rule, the variance in the data set was more than could have been attributed to sampling error alone. Forty-seven (47%) of the variance in the data set was the variance between outcome subgroups, indicating that more variance existed than attributed to sample artifacts. Further explanation regarding potential moderators in SGPP studies is provided later in this chapter.

## **Moderator Analysis**

To determine if moderators were present, outcome categories for each of the three meta-analyses (i.e., POWC, PPWC, and SGPP) were grouped by five potential moderating variables identified from study characteristics in the meta-analysis sample: training content, organization type, job classification, publication type, and measurement method. An analogue weighted ANOVA for each potential moderator indicated that organization type, job classification, and measurement method possibly impacted the results in certain outcome categories of POWC and SGPP studies. Table 19 reflects potential moderating variables explored in each outcome subgroup after adjustments for cells with only one study. It was interesting to note that no moderating variables appeared in the PPWC data set.

Two additional moderators (i.e., design type and objectivity-subjectivity) were explored across all studies in this meta-analytic research sample. These two moderators are discussed separately below as they were explored with a data set of studies from all three meta-analyses (POWC, PPWC, and SGPP). Thus, the implication of a design type or objectivity-subjectivity moderator, if found, was not necessarily related to a specific outcome subgroup within any of the individual meta-analyses.

The moderator results described below should be viewed with caution for two reasons: the low power of the studies and the probability of experiment-wise error. First, in many of the moderator-outcome combinations in this meta-analysis, there were subgroups analyzed that had a very small number of effect sizes. In many subgroups, only one or two effect sizes existed, which may mean that the power was too low to detect all moderator effects in those combinations. In addition, small numbers of studies

generate a distribution of effect sizes, which is what is expected in effective meta-analytic research. Thus, any one of the moderator effects presented in this research could possibly be an artifact due to the small number of effect sizes in sub-groups of the moderator variable.

Table 19: Variances Not Explained by Sampling Error in Moderator Analyses per Outcome Subgroup

	<b>POWC</b>	<b>PPWC</b>	<b>SGPP</b>
<b>EXPERTISE OBJECTIVE</b>	<b>(ES .329)</b>	<b>(ES .322)</b>	<b>(ES 1.01)</b>
Training Content	4%	8.2%	4.75%
Organization Type	33.9% <sup>a</sup>	19.8%	88.0% <sup>a</sup>
Job Classification	12.0%	19.9%	33.8% <sup>a</sup>
Measurement Method	.8%	5.7%	19.8%
Publication Type	1.1%	9.8%	.01%
<b>EXPERTISE SUBJECTIVE</b>	<b>(ES .30)</b>	<b>(ES .405)</b>	<b>(ES .378)</b>
Training Content	5.4%	11.6%	.05%
Organization Type	17.7%	27.6%	19.6%
Job Classification	9.5%	11.6%	.03%
Measurement Method	12.9%	.01%	27.8% <sup>a</sup>
Publication Type	1.9%	11.1%	.04%
<b>KNOWLEDGE OBJECTIVE</b>	<b>(ES .963)</b>		<b>(ES 1.37)</b>
Training Content	9.7%		15.0%
Organization Type	7.2%		.01%
Job Classification	38.4% <sup>a</sup>		8.7%
Measurement Method	.5%		.04%
Publication Type	2.5%		18.8%
<b>SYSTEM OBJECTIVE</b>	<b>(ES .390)</b>		
Training Content	.1%		
Organization Type	13.2%		
Job Classification	15.1%		
Measurement Method	0		
Publication Type	0		

Note. <sup>a</sup>Adjusted between-group variance greater than 25% reflecting possible moderator

Second, with a 5% probability of finding a moderator in 45 moderator-dependent variable combinations, it is likely that three cell combinations would have moderators. Five cell combinations in this meta-analytic research were found to have moderators.

Therefore, one could expect that approximately half of the moderators would occur through random sampling, making the overall impact of this moderator analysis even more suspect. Thus, the information on moderator analysis in this meta-analytic research is presented solely as an opportunity for future research. The findings should be interpreted with caution.

### **POWC Study Moderators**

The POWC design studies were grouped by each of the four outcome categories (i.e., expertise-objective, expertise-subjective, knowledge-objective and system-objective), for each of the potential moderating variables: training content, organization type, job classification, publication type, and measurement method. According to Hunter and Schmidt (1990), when between group variance is greater than 25% the results indicated that the variance is due to moderating variables. After combinations were discarded where there was only one study or two effect sizes by one author, the between group variance was reduced in most cells. The presence of possible moderators appeared in the following dependent variable-moderator combinations: expertise-objective/organization type and knowledge-objective/job classification (See Table 19).

Organization Type as a Moderator (POWC Expertise-Objective Outcomes). An analogue weighted ANOVA was run on the POWC expertise-objective studies and the between group variance for organization type was 44.6%. Table 20 shows that the initial  $Q_B$  (42.16) was 44.6% of  $Q_T$  (94.48).

In a second analog of ANOVA as shown in Table 21, where categories with a small number of studies, or all studies by the same author, were disregarded, the  $Q_B$

Table 20: Effect of Organization Type as a Moderator in POWC Design Studies with Expertise-Objective Outcomes

						95% Confidence Interval		
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit Upper Limit
Between groups	Q <sub>B</sub>	42.16	6					
Within groups	Q <sub>W</sub>	52.32	8					
Education	Q <sub>W1</sub>	.00	0	1	23	.73	.46	-.16 1.77
Government	Q <sub>W2</sub>	1.88	1	2	59	1.07	.28	.49 1.73
Manufacturing	Q <sub>W3</sub>	1.31	2	3	60	1.47	.28	.84 2.22
Medical	Q <sub>W4</sub>	19.13	1	2	79	.98	.24	.49 1.53
Military	Q <sub>W5</sub>	2.49	1	2	30	-.59	.42	-1.49 .22
Other	Q <sub>W6</sub>	26.38	2	3	233	.04	.13	.30 .41
Technology	Q <sub>W7</sub>	1.13	1	2	221	.14	.14	.12 .49
Total	Q <sub>T</sub>	94.48	14	15	705			

(26.17) was 33.9% of the total variance ( $Q_T=77.17$ ). This finding indicated that organization type remained as a possible moderating variable. Caution is advised in interpreting these findings for the following reasons: a) a wide range in the effect sizes (-.59 to 1.47) and b) the small number of studies in each cell. Two of the cells had only two studies and three cells had three studies. Further research should be conducted specifically relating to the impact that organization type plays upon the effectiveness of managerial leadership development programs.

Table 21: Adjusted Effect of Organization Type as a Moderator in POWC Design Studies with Expertise-Objective Outcomes

						95% Confidence Interval		
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit Upper Limit
Between groups	Q <sub>B</sub>	26.17	4					
Within groups	Q <sub>W</sub>	51.01	6					
Government	Q <sub>W1</sub>	1.88	1	2	59	1.07	.28	.49 1.73
Medical	Q <sub>W2</sub>	19.13	1	2	79	.98	.24	.49 1.53
Military	Q <sub>W3</sub>	2.49	1	2	30	-.59	.42	-1.49 .22
Other	Q <sub>W4</sub>	26.38	2	3	233	-.04	.13	-.30 .23
Technology	Q <sub>W5</sub>	1.13	1	2	221	.14	.14	-.12 .41
Total	Q <sub>T</sub>	77.17	10	11	622			

Job Classification as a Moderator (POWC Knowledge-Objective Outcomes). An analogue weighted ANOVA was produced for each moderator grouping for POWC knowledge-objective studies. The analogue weighted ANOVA for job classification indicated that 38.4% of the variance was related to the variance between job classification variables. Table 22 shows that the value for  $Q_B$  of 7.40 was 38.4% of  $Q_T$  (19.29). Because the variance between groups was greater than 25% (Hunter & Schmidt, 1990), the variance in the job classification variable was considered to be more than would be expected from sampling error alone. Thus, the findings indicated that job classification was a possible moderator in POWC design studies with knowledge-objective outcomes.

However, it should be noted that four effect sizes appeared in the mid-manager and supervisor-foreman subgroups and five in the mixed subgroup. Again, caution is advised in interpreting this finding because of the small number of studies in each subgroup. Therefore, this finding deserves further research to be conclusive.

Table 22: Effect of Job Classification as a Moderator in POWC Design Studies with Knowledge-Objective Outcomes

							95% Confidence Interval		
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit	Upper Limit
Between groups	Q <sub>B</sub>	7.40	2						
Within groups	Q <sub>W</sub>	11.89	10						
Mid-Manager	Q <sub>W1</sub>	4.01	3	4	284	1.25	.12	.98	1.54
Mixed	Q <sub>W2</sub>	7.84	4	5	442	.88	.10	.68	1.09
Supervisor-Foreman	Q <sub>W3</sub>	.05	3	4	151	.67	.17	.33	1.04
Total	Q <sub>T</sub>	19.29	12	13	877				



### **PPWC Study Moderators**

An analogue weighted ANOVA was produced for the two outcome subgroups (i.e., expertise-objective and expertise-subjective) for each potential moderator variable in PPWC studies: training content, organization type, job classification, publication type, and measurement method. Therefore, 10 outcome-moderator combinations were analyzed. The organization type subgroup in expertise-objective studies and organization type and training content subgroups in expertise-objective studies were explored in depth as moderator variable because unadjusted variance in the data set was initially greater than 25% (Hunter & Schmidt, 1990). However, after disregarding combinations with a small number of studies, the between group variance was reduced below the 25% level in each of the dependent variable-moderator combinations. These findings indicated that no apparent moderator variables existed in the PPWC sample in this meta-analysis, as the variance for these variables was no more than would occur through sampling error alone.

### **SGPP Study Moderators**

An analogue weighted ANOVA was produced for expertise-objective, expertise-subjective, and knowledge-objective SGPP research design studies for each potential moderator: training content, organization type, job classification, publication type, and measurement method. Therefore, 15 possible outcome-moderator combinations were analyzed. The organization type and job classification in expertise-objective studies and measurement method in expertise-subjective studies were explored in depth as moderator variables because unadjusted variance in the data sets was initially greater than 25% (Hunter & Schmidt, 1990). After disregarding categories where there was only one study or two effect sizes by one author, the between group variance was reduced, but the

presence of moderators appeared to remain (See Table 19). Those instances where moderators appeared to be present are discussed in detail in the following sections.

Organization Type as a Moderator (SGPP Expertise-Objective Outcomes). The between group variance for SGPP expertise-objective organization type studies ( $Q_B = 72.43$ ) as shown in Table 23 was 92.5% of the total variance ( $Q_T = 80.04$ ) within the data set in the initial analogue weighted ANOVA. This finding indicated that further exploration was needed to assess whether or not organization type was a potential moderator in the data set. Upon disregarding combinations with a small number of

Table 23: Effect of Organization Type as a Moderator in SGPP Design Studies with Expertise-Objective Outcomes

							95% Confidence Interval		
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit	Upper Limit
Between groups	Q <sub>B</sub>	72.43	4						
Within groups	Q <sub>W</sub>	7.61	9						
Education	Q <sub>W1</sub>	1.00	3	4	249	.17	.13	-.08	.43
Government	Q <sub>W2</sub>	.76	1	2	439	1.45	.10	1.22	1.68
Medical	Q <sub>W3</sub>	.00	1	2	22	.82	.52	-.18	2.02
Other	Q <sub>W4</sub>	5.84	4	5	255	1.36	.13	1.06	1.68
Technology	Q <sub>W5</sub>	.00	0	1	39	-.28	.34	-.96	.38
Total	Q <sub>T</sub>	80.04	13	14	1,004				

studies, the variance between organization types reduced from 92% to 88% (See Table 24). This finding indicated that organization type remained as a moderating variable. However, for the following reasons, this finding should be interpreted with caution: a) the wide range of effect sizes (from .17 to 1.45) and b) one cell had one effect size and two

Table 24: Adjusted Effect of Organization Type as a Moderator in SGPP Design Studies with Expertise-Objective Outcomes

								95% Confidence Interval	
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit	Upper Limit
Between groups	Q <sub>B</sub>	57.98	3						
Within groups	Q <sub>W</sub>	7.61	9						
Education	Q <sub>W1</sub>	1.00	3	4	249	.17	.13	-.08	.43
Government	Q <sub>W2</sub>	.76	1	2	439	1.45	.10	1.22	1.68
Medical	Q <sub>W3</sub>	.00	1	2	22	.82	.52	-.18	2.02
Other	Q <sub>W4</sub>	5.84	4	5	255	1.36	.13	1.06	1.68
Total	Q <sub>T</sub>	65.59	12	13	965				

cells had only two effect sizes. It should be noted that technology had one effect size, education and technology had two effect sizes each, education and government had four and other five. Of special note also in these results is that education as an industry had a low overall effect size of .17.

Job Classification as a Moderator (SGPP Expertise-Objective Outcomes). An analysis of job classification as a moderator in SGPP expertise-objective studies showed that 33.8% of the variance was between the job classification level of participants (See

Table 25: Effect of Job Classification as a Moderator in SGPP Design Studies with Expertise-Objective Outcomes

								95% Confidence Interval	
Source		Q value	Df	K	# cases	Effect Size	Standard Error	Lower Limit	Upper Limit
Between groups	Q <sub>B</sub>	27.06	3						
Within groups	Q <sub>W</sub>	52.97	10						
Mid-Manager	Q <sub>W1</sub>	.00	0	1	12	.85	.71	-.48	2.56
Mixed	Q <sub>W2</sub>	18.25	6	7	534	1.27	.09	1.07	1.48
Supervisor-Foreman	Q <sub>W3</sub>	32.24	2	3	331	.93	.11	.69	1.17
Top Management	Q <sub>W4</sub>	2.48	2	3	127	.15	.18	-.21	.51
Total	Q <sub>T</sub>	80.04	13	14	1,004				

Table 25). After a category with one study was discarded from the analysis, a second analogue weighted ANOVA showed that the between group variance still remained at 33.8% (see Table 26). Therefore, job classification appeared to be a possible moderator in the SGPP expertise-objective meta-analysis because the variance between interventions based upon job classification levels remained significantly greater than 25% (Hunter & Schmidt, 1990). However, this finding should be interpreted with caution because of the small sample size in each subgroup and the wide range of observed effect sizes (from .15 in the top management level and 1.27 for mixed levels of participants). The low power in this analysis could easily prevent other moderators from being detected. This finding presents opportunities for future research regarding moderating variables in SGPP studies, and specifically as gain scores are measured at the top management level.

Table 26: Adjusted Effect of Job Classification as a Moderator in SGPP Design Studies with Expertise-Objective Outcomes

								95% Confidence Interval	
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit	Upper Limit
Between groups	Q <sub>B</sub>	27.02	2						
Within groups	Q <sub>W</sub>	52.97	10						
Mixed	Q <sub>W1</sub>	18.25	6	7	534	1.27	.09	1.09	1.48
Supervisor-Foreman	Q <sub>W2</sub>	32.24	2	3	331	.93	.11	.69	1.17
Top Management	Q <sub>W3</sub>	2.48	2	3	127	.15	.18	-.21	.51
Total	Q <sub>T</sub>	79.99	12	13	992				

#### Measurement Method as a Moderator (SGPP Expertise-Subjective Outcomes).

The expertise-subjective subgroup data set of SGPP studies was grouped by each moderator variable of interest in the meta-analytic study. An analogue weighted ANOVA

by measurement method indicated that the variance between measurement method subgroups (self and other measurement) was 27.8%. Table 27 shows that a  $Q_B$  of 17.68 was 27.8% of  $Q_T$  (63.65). Therefore, measurement method appeared to be a possible moderator in this meta-analysis of SGPP expertise-subjective studies. It should be noted that only two effect sizes existed in the other category, and the finding deserves further research to be conclusive. Therefore, this finding should be used with caution.

Table 27: Effect of Measurement Method as a Moderator in SGPP Design Studies with Expertise-Subjective Outcomes

						95% Confidence Interval		
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit Upper Limit
Between groups	$Q_B$	17.68	1					
Within groups	$Q_W$	45.97	13					
Other	$Q_{W1}$	21.87	1	2	224	.95	.14	.67 1.26
Self	$Q_{W2}$	24.10	12	13	2,414	.33	.04	.25 .41
Total	$Q_T$	63.65	14	15	2,638			

### **Research Design as a Moderator**

In assessing the effectiveness of managerial leadership development programs, it was important to determine if the type of research design moderated the effect size of the studies in this meta-analytic research. There is much controversy in the literature regarding research design and the inclusion of studies without a control group. The decision was made to include SGPP studies in this research because single group pretest-posttest measurements are common in many training programs, and often the only evaluation method used. In addition, Hunter and Schmidt (1990) believe that the SGPP design to be far superior because it controls subject by treatment interaction.

The process of determining whether research design was a moderator was accomplished by running an analogue weighted ANOVA across the three meta-analyses: POWC, PPWC, and SGPP. The only two common outcome categories across all three meta-analyses were the expertise-objective and expertise subjective groups, and those were the studies used to determine whether or not research design was a moderator. However, the expertise-objective and the expertise-subjective data sets were analyzed separately for research design as a moderator as the two outcome categories primarily represent different measurement methods.

An analogue weighted ANOVA on all expertise-objective studies indicated that 25.49% of the variance in the data set was between research designs. As shown in Table 28, the  $Q_B$  of 68.34 was 25.49% of  $Q_T$  (268.08). Applying Hunter and Schmidt's (1990) 75% rule to this ANOVA, the between group variance for research design type was slightly greater than 25%, reflecting that the amount of variance was greater than that attributed to sampling error. Therefore, research design type appeared to be a possible moderator in expertise-objective studies in this meta-analytic research.

Table 28: Effect of Research Design as a Moderator in Managerial Leadership Development Interventions with Expertise-Objective Outcomes

		95% Confidence Interval							
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit	Upper Limit
Between groups	$Q_B$	68.34	2						
Within groups	$Q_W$	199.74	49						
POWC	$Q_{W1}$	94.48	14	15	705	.33	.08	.17	.49
PPWC	$Q_{W2}$	25.23	22	23	1,502	.32	.05	.22	.43
SGPP	$Q_{W3}$	80.04	13	14	1,004	1.00	.06	.87	1.15
Total	$Q_T$	268.08	51	52	3,211				

The results of an analogue weighted ANOVA for expertise-subjective studies shown in Table 29 indicated that the variance between research design types was only 2.08%. Therefore, in expertise-subjective studies in this meta-analysis, the research design type did not appear to be a moderator. Again, because of the small number of subjects within some subgroups (e.g., 469 in PPWC expertise-subjective), these findings should be used cautiously. These findings related to research design as a possible moderator presents opportunities for future research, especially because no known meta-analysis has been conducted using SGPP studies.

Table 29: Effect of Research Design as a Moderator in Managerial Leadership Development Interventions with Expertise-Subjective Outcomes

						95% Confidence Interval		
Source		Q value	Df	k	N	Effect Size	Standard Error	Lower Limit Upper Limit
Between groups	$Q_B$	2.54	2					
Within groups	$Q_W$	119.24	55					
POWC	$Q_{W1}$	36.46	23	24	1,335	.27	.06	.16 .38
PPWC	$Q_{W2}$	19.13	18	19	469	.40	.10	.20 .60
SGPP	$Q_{W3}$	63.65	14	15	2,638	.38	.04	.30 .46
Total	$Q_T$	121.78	57	58	4,442			

### **Objective versus Subjective Outcomes as a Moderator**

To determine whether objectivity-subjectivity was a moderator was accomplished by running an analogue weighted ANOVA across the three meta-analyses: POWC, PPWC, and SGPP (See Table 30). The only two common outcome categories across all three meta-analyses were the expertise-objective and expertise-subjective groups, and those were the studies used to comprise the data set for the objectivity-subjectivity moderator analysis.

The analogue weighted ANOVA of all expertise-objective and expertise-subjective studies shown in Table 30 indicated that 3.57% of the variance in the studies was related to the difference in objective versus subjective outcomes ( $Q_B = 14.28$  and  $Q_T = 404.14$ ). According to Hunter and Schmidt's (1990) 75% rule, the amount of variance appeared to be no greater than that attributed to sampling error. Therefore, objective or subjective outcomes did not appear to be moderating variables in this meta-analysis. However, further analysis should be done in this area prior to being confident in these findings.

Table 30: Effect of Objective-Subjective Outcomes as a Moderator in Managerial Leadership Development Interventions from 1982-2001

						95% Confidence Interval		
Source		Q value	Df	k	Subjects	Effect Size	Standard Error	Lower Limit Upper Limit
Between groups	$Q_B$	14.28	1					
Within groups	$Q_W$	389.86	108					
Expertise-Objective	$Q_{W1}$	268.08	51	52	3,211	.53	.04	.46 .61
Expertise-Subjective	$Q_{W2}$	121.78	57	58	4,452	.35	.03	.29 .41
Total	$Q_T$	404.14	109	110	7,653			

### Summary

This chapter described the meta-analysis data set, presented the findings of this research, and examined potential moderator variables. Eighty-three (83) studies on managerial leadership development from 1982-2001 were divided into three independent meta-analyses according to research design: posttest-only with control group (POWC), pretest-posttest with control group (PPWC), and single group pretest-posttest (SGPP). A



separate meta-analysis was performed on each data set. The overall effect size for each outcome subgroup is presented in Table 31.

Table 31: Summary of Effectiveness Levels of Managerial Leadership Development Interventions from 1982-2001 by Outcome Subgroups per Research Design

Outcome Subgroup	Design	Effectiveness Level			k	N	Moderator
		Low	Moderate	High			
Knowledge-Objective							
	POWC	-	-	.96 <sup>a</sup>	13	877	Job Class
	PPWC	-	-	-	-	-	-
	SGPP	-	-	1.37	6	642	-
Expertise-Objective							
	POWC	-	.33 <sup>a</sup>	-	15	705	Org Type
	PPWC	-	.32	-	23	1,502	-
	SGPP	-	-	1.01 <sup>a</sup>	14	1,004	Org Type Job Class
Expertise-Subjective							
	POWC	.30	-	-	24	1,335	-
	PPWC	-	.40	-	19	469	-
	SGPP	-	.38 <sup>a</sup>	-	15	2,638	Measurement Method
System-Objective							
	POWC	-	.39	-	7	418	-
	PPWC	-	-	-	-	-	-
	SGPP	-	-	-	-	-	-

Note. <sup>a</sup> Possible moderators present. A dash indicates that the mean effect size did not fall within the respective effectiveness level.

Overall, most managerial leadership development interventions in this meta-analysis were found to be from moderately to highly effective. Formal training programs with knowledge outcomes were highly effective. The average effect size for knowledge outcomes ranged from .96 (control group, knowledge-objective) to 1.37 (pretest-posttest, knowledge-objective). The average effect size for expertise outcomes ranged from .30 (control group, expertise-subjective) to 1.01 (pretest-posttest, expertise-

objective), and those with system outcomes had an average effect size of .39 (control group). This meta-analysis synthesized existing studies from a broad range of settings, researchers and circumstances and integrated conflicting findings to establish a general knowledge base about managerial leadership development.

## **CHAPTER 5: DISCUSSION AND CONCLUSIONS**

This meta-analysis aggregated the results from 83 studies on managerial leadership development outcomes from 1982-2001, with formal training interventions, published primarily in psychology and business/management sources. The majority of interventions had behavioral outcomes and a human relations or general management training content focus. In the studies there appeared to be a trend toward multiple training techniques, a blend of cognitive knowledge and behavioral learning, and multiple evaluation techniques that included evaluations by supervisors, subordinates, and peers, along with self-assessments.

This meta-analysis was undertaken to understand the magnitude of the effectiveness of managerial leadership development programs by outcome subcategories and research designs. However, it is important to note that a relatively small number of observed effect sizes were obtained for system-subjective, financial-objective, and knowledge-subjective outcomes, which prohibited the inclusion of these studies in this meta-analysis (See Tables 10 and 11 in Chapter 4). The studies included in the sample were divided into data sets according to the type of research design: posttest only with control group (POWC), pretest-posttest with control group (PPWC), and single group pretest-posttest (SGPP). Fifty-nine (59) effect sizes were aggregated for POWC studies, 42 for PPWC, and 35 for SGPP studies to determine the effectiveness of managerial leadership development programs.

This chapter discusses the findings and conclusions of this meta-analytic research and compares the results with previous meta-analyses. Also presented in this chapter are the limitations of this meta-analysis and implications on practice and future research.

## Research Questions

The conclusions of this meta-analytic research are presented based upon the five research questions in Chapter 1. For each of the four research design types (POWC, PPWC, SGPP, and CORR), the research questions are answered.

### **Research Question 1: Across Studies Measuring System Outcomes, How Effective Is Managerial Leadership Development?**

Each year government and private industry in the United States spend billions of dollars on managerial leadership development (Bassi, Benson, & Cheney, 1996; Gibler, Carter, & Goldsmith, 2000). In addition, organizations have been asked to become more accountable and justify the existence of programs in terms of costs and performance factors (Dionne, 1996). It is believed that increasing global competition has led to pressure to demonstrate that programs are contributing to the “bottom-line” of the organization (Holton, 1995). However, there is surprisingly little reported systematic evaluation of training programs with organizational performance as an outcome (Collins, 2001; Sogunro, 1997).

The current POWC research located only seven studies with system outcomes from 1982-2001. The effect size for system-objective outcomes in those studies was .39 with 418 subjects, with effect sizes ranging from .02 to .79. This means that the performance-level outcome measured objectively for the trained group was .39 standard deviation higher than the control group, indicating that interventions with system outcomes were moderately effective. Studies found by Burke and Day (1986) and Zhang (1999) contained interventions with system outcomes as well. Table 32 indicates that the effect size for system-objective outcomes in Burke and Day’s study was .67 (2,298 participants) and for Zhang’s study was .49 (392 participants).

Table 32: Comparison of Meta-Analyses on the Effectiveness of Managerial Leadership Development Programs by Outcome Subgroup

	Burke and Day 1986 70 Studies 472 Effect Sizes 3,967 Subjects	Zhang 1999 29 Studies 61 Effect Sizes 3,352 Subjects	Collins POWC 2002 36 studies 59 Effect Sizes (3,335 Subjects)	Collins PPWC 2002 26 studies 42 Effect Sizes (1,971 Subjects)	Collins SGPP 2002 25 studies 35 Effect Sizes (4,284 Subjects)
Outcome Subgroup					
System					
Objective	.67	.49	.39	-	-
Subjective	-	.25	-	-	-
Expertise					
Objective	-	-	.33 <sup>a</sup>	.32	1.01 <sup>a</sup>
Subjective	.49	.50	.30	.40	.38 <sup>a</sup>
Knowledge					
Objective	.38	.80	.96 <sup>a</sup>	-	1.37
Subjective	.34	.47	-	-	-

Note. <sup>a</sup> Possible moderators present. A dash indicates that an effect size was not determined for the outcome subgroup.

While these differences in effect sizes are notable, it is important to point out that there are methodological differences between the three meta-analyses. It is also important to note that these methodological differences give reason to be cautious about comparisons of effect sizes of the current meta-analysis with the other two.

First, Burke and Day (1986) calculated an effect size for each dependent variable within a single study. The unit of analysis on which the effect sizes were combined for the current meta-analysis was a single outcome measure for the study. When more than one dependent measure was used to test the same relationship in a single study, a weighted average effect size was calculated where possible to produce one effect size per outcome per study. To point out the impact, Burke and Day aggregated studies involving 3,967 treated and 3,186 control group participants, but reported 46,574 total subjects across 472 effect sizes. For comparison, this research reported 9,590 total subjects across 136 effect sizes in POWC, PPWC, and SGPP research designs.

Burke and Day's (1986) methodology raises two primary issues: 1) independence of outcomes measured (effect sizes), and 2) over-weighting of studies with multiple effect sizes. For example, an intervention with four subscales would generate four effect sizes in Burke and Day's meta-analysis, but would generate only one in this meta-analysis. In addition, these four effect sizes were not independent as required by meta-analysis assumptions because "any two or more effect sizes that come from the same subject sample are statistically dependent..." (Burke and Day's) procedure potentially introduces substantial error as the inflated sample size, the distortion of standard error estimates arising from the inclusion of nonindependent effect sizes, and the overrepresentation of those studies that contribute more effect sizes can render the statistical results highly suspect" (Lipsey & Wilson. 2001. p. 105).

A second difference is that Burke and Day (1986) combined all behavioral outcomes into a *subjective behavior* subgroup, whereas the current research delineated behavioral outcomes into two subgroups: *expertise-subjective* and *expertise-objective*. Also, results outcomes (both financial and system) were combined into *objective results* by Burke and Day. This meta-analysis did not mix the objective and subjective outcomes, nor did it mix financial and system outcomes. To clarify, where this meta-analysis described four outcome subgroups as *system-subjective*, *system-objective*, *financial-subjective*, and *financial-objective*, Burke and Day described them as a single *objective results* group. Thus, a problem exists of comparing "apples and oranges" when comparing effect sizes.

Combining the outcome subgroups as was done by Burke and Day (1986) is problematic, as it is unreasonable to believe that the standard error for financial returns

would be the same as the standard error for system outcomes. To combine outcome subgroups is to assume that the outcomes are equivalent with equivalent distributions. Defining the outcome subgroupings as in this meta-analysis is more advanced and refined, and believed to be a better approach. Actually, Burke and Day's methodology would not be used in current practices as meta-analytic procedures have improved during the last 20 years. If Burke and Day's study were conducted with newer practices (e.g., newer formulas and definitions) the value obtained for the impact of managerial training would most likely be different.

A third pertinent difference is that Burke and Day (1986) used studies from a business and industry setting only, whereas the current analysis aggregated findings from education, government, medical, military, and others in addition to business and industry. Therefore, effect sizes would likely be different for that reason alone.

Of key importance to the comparison of effect sizes at the system-objective level is the low number of studies (two studies) in Burke and Day's (1986) meta-analysis. It should be noted that 11 studies in this overall meta-analysis research had organizational-level outcomes. This raises a concern of how representative Burke and Day's system-objective effect size would be, and also causes concern in the comparison of findings. Managerial leadership development is a young field for which little information is reported in the literature regarding what is or what is not effective, particularly relative to outcomes of the organization as a system.

Zhang's (1999) meta-analysis more closely resembles the current meta-analysis. However, she combined all behavioral outcomes into a *subjective behavior* subgroup, and all financial results were combined with system outcomes in an effort to replicate Burke

and Day (1986). To clarify, this meta-analysis did not mix the objective and subjective outcomes, nor did it mix financial and system outcomes. For instance, Zhang defined all behavior results as subjective behavior while this meta-analysis defined the same results as *expertise-objective* and *expertise-subjective*. Whereas this meta-analysis described performance outcome subgroups as *system-subjective*, *system-objective*, *financial-subjective*, and *financial-objective*, Zhang described them as two groups: *subjective results* and *objective results* (system-subjective and system-objective). Thus, to contrast effect sizes from the studies would again be like comparing “apples and oranges”, and this is problematic, as it assumes that the outcomes are equivalent with equivalent distribution. Definitions of outcome subgroupings as in this meta-analysis are believed to be the best approach.

It is also important to note that this meta-analysis located 49 studies, which would have met Zhang’s (1999) criteria for inclusion, contrasted to 29 studies located by Zhang. This leads to a question about the soundness of Zhang’s literature search and the possible generalization of Zhang’s findings. It is interesting to note that Zhang’s (1999) studies were from mixed organization types, primarily from business and industry and education.

Studies with system outcomes varied greatly in this meta-analysis. Some examples of studies with system outcomes are as follows:

- 1) Student achievement scores and community relations were an outcome of an in-service program for principals (Bankston, 1993). The effect size was .79 indicating that scores for the trained group scores were .79 standard deviation greater than the control group.



- 2) Number of referrals to employee assistance programs was an outcome of training for entry-level supervisors (Colan & Schneider, 1992). This intervention was minimally effective with an effect size of .30.
- 3) Detection of production errors after training was the outcome of an intervention for a mixed group of managers (Graen, Novak, & Sommerkamp, 1982). The effect size was .60 indicating that the detection of errors by the trained managers was .60 standard deviation greater than detection by the control group.
- 4) Change in organizational culture was a result of a leadership development model for top managers, measured by a Culture Analysis Questionnaire by subordinates (Hill, 1992). This intervention was minimally effective with an effect size of .10.

Unfortunately, this meta-analysis showed that there is little research describing strategic stewardship training programs, or system-level outcomes that involve transformational leadership primarily at the top management level. One can speculate that it is too soon for significant research to occur in the literature regarding strategic stewardship and the need to train leaders in strategy development. However, as a result of this meta-analysis it is realized that more research should be done in this area to determine if organizations are focusing on performance at the system level, or to see if this type of training is working.

**Research Question #2: Across Studies Measuring Financial Outcomes, How Effective is Managerial Leadership Development?**

There is a tremendous deficiency in the research on managerial leadership development programs regarding financial outcomes. More research is needed, as the

inability to conduct a meta-analysis on the financial outcome subgroup is indicative of missing research. Therefore, the effectiveness of managerial leadership development programs across studies measuring financial outcomes could not be estimated, and conclusions cannot be drawn regarding financial outcomes until adequate empirical studies are performed.

Few studies are available perhaps because financial performance (or overall profitability) would be less responsive to individual behavior change in the short time period typically needed to train individuals, evaluate the training program, and report the results in the literature. Evaluations of programs with a financial outcome would require longer periods of time than many companies are willing to devote. In addition, organizations are typically resistant to publishing financial outcomes as a result of training programs, especially when the results are negative. Therefore, organizations are more likely to measure knowledge or behavior outcomes that are thought to be responsive to leaders' behaviors within the time frame of the study.

In the literature search for this meta-analysis, only two studies with financial outcomes were located (Barling, Weber, & Kelloway, 1996; Krug, 1992). Those two studies will be described to provide a greater understanding of what interventions are being reported in the literature with financial outcomes.

Barling, Weber, and Kelloway (1996) found that transformational leadership training was effective for two aspects of branch-level financial performance. The criteria for measurement of effectiveness of the intervention was the number of personal loan sales and number of credit card sales taken from the region's regular records. These two variables were chosen as they were expected to be responsive to branch manager's

transformational leadership, which would presumably raise employee expectations, clarify the mission and challenge assumptions about unproductive performance methods. This intervention was found to be moderately effective with a .42 effect size, which means that the trained group of branch managers performed .42 standard deviation higher than the control group.

Krug (1992) examined the cost effectiveness of a model in-service program for developing effective instructional leaders. The intervention was built on the belief that effective leadership for schools was dependent upon the role of the principal. The intervention involved one-on-one interaction between the principal and a “leadership analyst” in the interpretation of assessment results. Krug used Hunter and Schmidt ‘s (1983) method of computing the standard deviation of job performance in dollars for the typical job in the U.S. economy (i.e., the standard deviation of output in dollar terms is approximately 40% of the average annual wage). Krug found that the in-service training program for principals was moderately effective (.37 effect size).

### **Research Question #3: Across Studies Measuring Expertise Outcomes, How Effective Is Managerial Leadership Development?**

Systematic evaluation of training programs should include the impact of training upon changes in work behavior (Alliger & Janak, 1989; Faerman & Ban, 1993), as without this kind of pertinent information, managers have a limited understanding of training and make costly decisions based upon reaction-level information only. Thus, training outcomes should emphasize individual (and organizational) performance and not just learning. This meta-analysis indicates that trainers are obviously making efforts to conduct evaluation beyond the reaction and learning levels to assess the performance of leaders and managers on the job.

Multiple performance appraisal instruments are prevalent in many organizations, with evaluations conducted by supervisors, subordinates, peers, and self-evaluations. Shipper and Neck (1990) believed that subordinates' observations were the most effective technique to pinpoint needed changes for individual performance improvement. It was interesting to note that 80% of the studies in this research had behavioral outcomes: expertise-objective, 38% and expertise-subjective, 42%. Interventions with expertise-objective outcomes were primarily measured by other person measurements, such as subordinates' or supervisors' evaluations. Expertise-subjective outcomes were primarily measured by self-assessments.

Expertise-Objective Outcomes. From this meta-analysis it appears that there is a trend to more research in programs with behavior outcomes, specifically those that are measured objectively. For instance, 80% of the studies in this meta-analysis measured managerial behavior outcomes as compared to 59% in Burke and Day's (1986) study and 39% in Zhang's (1999) study. It is important to note that both Zhang's and Burke and Day's meta-analyses combined subjective and objective behavior into the expertise-subjective outcome subcategory to report findings. It is also important to remember that Zhang's study does not represent all findings in the literature on managerial leadership development, and thus some behavioral studies are missing. Finally, it is important to note in the current meta-analysis that findings with expertise-objective outcomes were located in all three study designs (i.e., POWC, PPWC, and SGPP).

Expertise-objective outcomes were found to be moderately effective across the POWC and PPWC meta-analyses, and highly effective in SGPP measurements. The overall effect size for POWC expertise-objective studies was .33 (aggregated from 15

effect sizes and 705 subjects). This finding indicates that the difference in behavior between the trained group and the untrained comparison group was .33 standard deviation. The overall effect size of PPWC expertise-objective studies was .32 (aggregated from 23 effect sizes and 1,502 subjects), an effect size almost identical to POWC studies. However, for SGPP studies, the effect size for expertise-objective outcomes was 1.01 (aggregating from 14 effect sizes and 1,004 subjects). It is important to note that the POWC data set varied greatly with a range of effect sizes from -1.39 to 1.99, the PPWC data set ranged from -.45 to 1.22, and SGPP from -.28 to 1.66.

The most obvious measure of behavioral change (expertise) is gain scores. A gain score is the difference between the measure of an individual's performance before training (pretest score) and the measure of performance after completion of the training (posttest score) -- as represented in SGPP studies. From the results of the current meta-analysis, it can be concluded that behavior change when measured objectively from pretest to posttest was approximately .7 standard deviation greater than when comparing the scores of a treatment and control group after an intervention.

Unfortunately, SGPP studies are often overlooked as a valuable resource in the effectiveness of managerial leadership development programs. Actually, Zhang (1999) excluded SGPP studies because she believed that SGPP "results had little research value" (p. 104). That statement seems odd, especially because single group pretest design is often the only type of evaluation design that can be used in certain training settings. In addition, Hunter and Schmidt (1990) "urged experimenters to use the more powerful within-subjects design whenever possible" (p. 340). Evaluations of behavioral change require a systematic assessment of job performance both before and after completion of

the intervention. Therefore, it is becoming more important for studies to be performed with pretest-posttest research design.

Nevertheless, researchers do not typically use SGPP, as they believe that the design is flawed by threats to internal and external validity. However, Carlson and Schmidt (1999) and Hunter and Schmidt (1990) demonstrate that this is not true. To control for some of those threats, Carlson and Schmidt's formulas, using pretest standard deviation instead of posttest standard deviation, were used in the computation of SGPP effect sizes in this meta-analysis. And, a global correction was made for error of measurement using reliability information from 46 studies. Thus, this meta-analysis incorporated adjustments that were believed to reduce the effect size inflation often seen in SGPP studies. Nonetheless, the effect size was still substantially higher.

In addition, adult learning principles alert us to the fact that individuals react differently to training based on their individual differences, a concept known as treatment by subject interaction. Basically, if the treatment has different effects on different people because of individual differences, then there is an interaction between treatment and subjects. It is no surprise that there is a wide range of individual differences among managers who participate in training programs, and for those training programs to be effective, they must accommodate individual managers' abilities, learning styles, and preferences. For example, some people learn best from lectures, others from structured exercises or direct experiences.

The pertinent point here is that pretest-posttest designs are the only ones that can incorporate the effect of treatment by subject interaction. Control group designs do not provide the data to capture the effect of treatment by subject interaction. Hunter and

Schmidt (1990) believe that “independent groups design is one in which half the data is missing” (p. 345). Or, as they continue, “all the data on individual treatment effects is missing.” Specifically, in POWC studies, the effect is assumed to be the same for every subject, which is not the case in most training programs.

To explain further, in POWC studies the difference between the group means is the treatment effect. It is merely assumed there is no treatment by subject interaction. For example, if the outcome for half the treated subjects is +5 and the other half is -5, the average treatment effect is zero, so the mean of the treatment group equals the mean of the control group. The logical interpretation is that the treatment had no effect, an interpretation that is totally false. Thus, when there is treatment by subject interaction, the POWC design is questionable. Therefore, the SGPP effect size for expertise-objective outcomes (1.01) may be the most reflective of the true effect size since treatment by subject interaction is incorporated.

The larger effect size for SGPP studies is not unusual. Leddick (1987) conducted a meta-analysis of training effectiveness using both SGPP and POWC studies, and found that “effects were smaller when true controls or non equivalent control groups were used” (p. 98). It was also interesting that Leddick obtained an effect size of .98 for a mixed group of expertise- subjective and objective outcomes, very close to the effect size obtained in this meta-analysis for the SGPP expertise-objective outcomes (1.01). Chen (1994) also found that in a meta-analysis on the effectiveness of cross-cultural training for managers that effect sizes from SGPP studies overall were higher than those with control groups (1.74 versus 1.58).

Because previous meta-analyses combined behavioral studies that are measured objectively into one effect size (subjective behavior, or expertise-subjective in this research), this research provides the only known pure research regarding expertise-objective findings. Combining the two outcome categories into one as was done by Zhang (1999) and Burke and Day (1986) is unwise, as subjective and objective results are distinct and reflect different measurement strategies. This is apparent by the amount of literature on self versus other measurement, which indicates that self-measurements are usually higher. Therefore, it is strongly believed that findings in this meta-analysis provide a greater understanding about managerial leadership development in terms of behavioral outcomes of training than any other meta-analysis.

Expertise-Subjective Outcomes. Interventions with expertise-subjective outcomes were measured by subjective methods, primarily self-assessment, and were found in studies from all three research designs (i.e. POWC, PPWC, and SGPP). Expertise-subjective outcomes were found to have a low effectiveness in POWC studies, and moderately effective in PPWC and SGPP studies. In the POWC meta-analysis, expertise-subjective outcomes (24 effect sizes) produced an average effect size of .30 from 1,335 subjects. This finding indicates that the training outcome for the trained group was .30 standard deviation higher than the control group. The observed effect size for PPWC expertise-subjective studies was .40 determined from 19 effect sizes and 469 subjects. This finding indicates that the difference in behavior change as determined by the trained group is .40 standard deviation higher than the untrained group. The effect size for expertise-subjective outcomes in SGPP studies was .38, aggregated over 15 effect sizes and 2,638 subjects. This means that individual participants in training programs indicated



that their ability to perform was .38 standard deviation greater after training than prior to training. It must also be pointed out that the SGPP expertise-subjective outcomes studies reported no negative finding as there was positive behavior change reported by self-assessments in all studies as a result of the intervention.

Burke and Day (1996) produced an effect size of .49 for expertise-subjective studies, and Zhang (1999) an effect size of .50. While this POWC meta-analysis produced an effect size of .30 (See Table 32 above), if expertise-objective and expertise-subjective data were combined into one for comparison purposes, the effect size would be .32. Overall expertise outcomes were moderately effective in all three meta-analyses, but the effect sizes were noticeably different. In addition, it is interesting to compare the effect size for expertise-subjective outcomes in the POWC meta-analysis to the effect size for POWC expertise-objective outcomes, and they are noticeably similar (.30 and .33 respectively).

The difference in effect sizes between this meta-analysis and the other two is very curious. One would expect some differences because Burke and Day (1986) limited their research to business and industry, which caused the overall focus to be somewhat different. Eighty three percent (83%), or 24 out of 29, studies by Zhang (1999) were from business and industry and education, compared to 41% (34 studies) of the current meta-analysis. This difference in composition by organization type, and an inadequate literature search by Zhang, suggests that a difference in the sample by organization type was likely the cause for such a wide difference in effect sizes.

Also, it is important to compare the expertise-subjective effect size of .30 in POWC studies with the effect size from the same outcome subgroups in PPWC studies

(.40) and in SGPP studies (.38) (See Table 32 above). That SGPP findings are higher than POWC is not surprising because it may be a stronger design that detects and measures treatment by subject interaction as discussed earlier (Hunter & Schmidt, 1990). Thus, the effect size of .38 for SGPP expertise-subjective studies in this meta-analysis may be the best representation of the true effectiveness of managerial leadership development with subjective behavior outcomes.

One other interesting finding regarding the measurement of behavior was that the effect size for objective behavior measurements was significantly higher than subjective behavior effect sizes in SGPP studies (1.01 objective versus .38 subjective). Objective measurements were primarily other people rating the participant's behavior after training. This finding was surprising because subjective ratings are usually higher than objective ratings. It is possible that self-raters do not see change in themselves as quickly as it detected by the supervisors or subordinates. In addition, some people are overly critical of themselves, and may not rate themselves as high as others would do.

#### **Research Question #4: Across Studies Measuring Knowledge Outcomes, How Effective Is Managerial Leadership Development?**

This research indicated that learning outcomes remain a focus of managerial leadership development programs and that interventions with knowledge outcomes are highly effective. It would stand to reason that managerial ranks would know why they needed the information provided in training, and could understand why it would be of benefit to them in their own positions. Knowledge-objective outcomes were meta-analyzed only in POWC and SGPP studies. It is important to remember that the few knowledge-objective outcomes in PPWC were combined with POWC in this meta-analysis (See Tables 10 and 11 in Chapter 4).

In POWC studies, the knowledge-objective overall effect size was .96, aggregated from 13 effect sizes and 877 subjects. This means that the training outcome measured primarily by knowledge tests of the trained group was almost one standard deviation higher than the untrained control group in POWC studies. This result compares with .38 effect size found by Burke and Day (1986) and .80 with Zhang (1999). Therefore, knowledge outcomes were found to be highly effective in this meta-analysis and in Zhang's study, while Burke and Day found them to be moderately effective. It was interesting to note that the effect sizes in this POWC meta-analysis ranged from .32 to 1.94.

Knowledge-objective results were also found in SGPP studies with an overall average effect of 1.37 from six effect sizes and 642 subjects. While the latter finding is significantly greater, it should be noted that there were a limited number of studies from which to base the determination. However, it was interesting that the effect sizes in SGPP studies ranged from .66 to 1.65 in SGPP studies and both the high and low effect size was in the supervisor-foreman job classification level.

The magnitude of the findings in the SGPP meta-analysis (1.37) regarding knowledge outcomes is not surprising. It would be logical to assume that in SGPP studies where individual differences are recognized (i.e. where treatment by subject interaction is incorporated and measured), the effect size would be higher. However, this area could use further research and especially with SGPP studies, in regard to treatment by subject interaction.

**Research Question #5: What Moderator Effects Can Be Detected For The Following Variables: Training Content, Organization Type, Job Classification Level, Publication Type, Measurement Method, Research Design, And Objective-Subjective Outcomes?**

The impact of five moderators was pursued in each of the meta-analyses (POWC, PPWC, and SGPP): training content, organization type, job classification level, publication type, and measurement method. In addition, subjectivity-objectivity and research design were tested as moderating variables across all studies with expertise outcomes. Because possible moderators were chosen from study characteristics, it is important to note that they were not an exhaustive list of possible moderators in managerial leadership development.

Caution was expressed throughout the findings regarding the soundness of the detection of moderators. In many circumstances, moderators were analyzed across subgroups with only two or three effect sizes. Variables found as possible moderators for selected outcome categories were reported in the findings in Chapter 4: organization type, job classification level, measurement method, and research design. Rather than offer a definitive interpretation, these moderators were suggested as areas of future research for the following reasons:

- 1) Low power that may have prevented the detection of other moderators in the outcome subcategory.
- 2) Low number of studies that may have skewed the data. A low number of studies with spurious findings, or with a unique measurement instrument, may have led to a wide range of effect sizes within subgroups and a false detection of the variable as a moderator.

- 3) The chance of experiment-wise error rate where one would expect moderators in 5% of the cell-moderator combinations.

One interesting finding was that self versus other measurement did not appear as a moderator in eight out of nine research design-outcome categories. Self-measurements were typically subjective measurements and other measurements were typically objective measurements. The literature usually reports that self-other measurements are significantly different, with self-assessments normally having higher reported results. Regardless, Shipper and Neck (1990) believe that self-observations tend to be biased, as they do not provide new information to the manager. One interpretation is that self ratings are not biased when measuring leadership development outcomes. However, in outcome subcategories where both self and other measurements appeared, either self or other heavily dominated that subcategory. For example, it was not uncommon to find 12 self-measurements and two other measurements in a subjective outcome subcategory. Therefore, low power or an occasional spurious result likely influenced this finding.

Three of the five possible moderators that were detected occurred in SGPP studies. A key finding in the SGPP expertise-objective moderator analysis was that the between group variance by organization type remained at 88% after small numbers of studies were disregarded. This indicated that more variance between variables existed in the data set than could be contributed to sample artifacts. It is important to note the effect sizes after subgroups with one study (or one author) were disregarded ranged from .06 to 1.71. This wide variance was across four organization types with two organizations having two studies each and the remaining two organizations having four and five studies each. It is important to note that two subgroups (government and a group where

organization type was not reported) had average effect sizes above 1.0. This moderator seems most likely to be a true moderator, and reflects an opportunity for an important area of further study.

Job classification also appeared as a moderator in SGPP studies with expertise-objective outcomes. What is curious about this finding is that the interventions for entry level managers had an effect size of .93 and top management an effect size of .14. It may be that entry-level managers have a greater need, or more motivation to learn, as they could potentially have more to gain from training than top management.

The publication type and training content did not appear to make a difference on the magnitude of the training effect. There was no a priori reason to anticipate that either would be a moderator in this meta-analysis. However, it is important to provide the right training content to the right group of people, and managers need to learn different content as they progress to another job classification level. When the training program is effectively done, it can be equally as effective for all types of content.

The type of research design appears to be a moderator variable for expertise-objective outcomes, but not for expertise-subjective outcomes. One would anticipate research design to be a very strong moderator in managerial leadership development across both expertise-objective and expertise-subjective studies for the following reasons:

- 1) SGPP studies are believed to be more powerful than other designs (Hunter & Schmidt, 1990).
- 2) SGPP design captures treatment by subject interaction (Hunter & Schmidt, 1990).
- 3) SGPP produces higher effect sizes (Chen, 1994; Leddik, 1987; Zhang, 1999).

Further research should be conducted on research design as a moderator, and specifically in relation to SGPP as a research design and the ability to capture treatment by subject interaction. A future research suggestion is to split PPWC studies in this meta-analysis into two groups and then conduct a moderator analysis to determine the impact of the design type.

Another interesting finding was that no moderators appeared to exist in the PPWC meta-analysis. There is no reason from the literature to believe that PPWC as a research design would not have moderators. Therefore, caution is advised regarding these findings until further research is conducted. According to Carlson and Schmidt (1999), PPWC studies contain more information – the pretest-posttest comparison plus the control-experimental group comparison – and it is logical to assume that possible moderators should have been detected in this data set, if they did indeed exist. However, as stated previously, the PPWC data set provides an opportunity for future research in that it could be split into SGPP and POWC and be analyzed separately. To do this might provide a better insight into the PPWC moderator analysis.

All moderators explored in this meta-analysis should be approached with caution. Therefore, until more research is done on the effectiveness of managerial leadership development, HRD professionals should not make definitive conclusions about moderators from this study.

### **Limitations of the Study**

Despite the strength of this study, certain limitations to this meta-analytic research should be noted:

- 1) The results are only as good as the studies included in the sample (Lipsey & Wilson, 2001). Some studies employed sophisticated designs that controlled for variables that likely affected their outcomes and other studies provided few controls. In an effort to be inclusive, all applicable studies were used for this meta-analysis.
- 2) Uneven reporting practices among researchers caused concern. For example, some studies that should have been reported in this research are missing because of insufficient statistical analysis.
- 3) This meta-analysis, as in the case of all meta-analytic research, required judgments in many areas leading to potential interpretation errors (Wanous, Sullivan, & Malinak, 1989).
- 4) The number of studies that met the criteria for the research was relatively small, especially those with system and financial outcomes. Therefore, all objectives of this research were not accomplished and moderator detection was inconclusive because of the low number of studies in various subgroups.
- 5) Some difficult-to-find studies were not included, particularly unpublished government documents and conference proceedings. However, it should be noted that a large number of unpublished studies (primarily dissertations) were incorporated.
- 6) Fifteen (15) studies out of 83 studies (26 effect sizes) only reported p-values from which to calculate effect sizes, which were not as precise as using means and standard deviations.



- 7) In many studies, the participants were aware that their behavior was under scrutiny.

### **Implications for Practice**

Most organizations sponsor leadership development programs for their managers and assume that such investments of time produce results. This research shows that organizations should feel comfortable that their managerial leadership development programs will produce substantial results, especially if they do the right development programs for the right people at the right time. For example, it is important to know if a six week training session is enough or the right approach to develop new competencies that change managerial behaviors, or is it individual feedback from a supervisor on a weekly basis regarding job performance that is most effective?

This meta-analysis indicates that a wide variety of formal training programs are occurring in organizations. But, it also shows that there is a wide variance in the effectiveness of those programs. This means that there are some tremendous programs, but some are failing miserably. However, this is not surprising, especially because there is not one clear, concise, ubiquitous definition of leadership, and because leadership is complex and quite difficult to measure objectively. As a result, leadership development programs may incorporate leadership dimensions in the program design that are not appropriate for the organization.

Nevertheless, the overall aggregated effect sizes in this meta-analysis are comforting. Training programs with expertise and system outcomes are effective, but most importantly, we can conclude that interventions that are intended to impart knowledge to the participants are highly effective and worthwhile. Therefore, the

potential for gain from a managerial leadership development program is substantial, especially if it is the right program for the organization. However, the burden is on HRD professionals to determine the appropriate intervention for the organization and provide good leadership development experiences that produce the kinds of leaders that are needed to meet the strategic goals of the organization. Some training professionals have made great efforts conducting training needs assessments and creating favorable conditions for transfer of training.

It is quite possible that one explanation for the wide variation in effectiveness reported is due to poor needs analysis. To be effective, training interventions should be preceded by an assessment of the organization's learning needs. A needs analysis helps ensure that resources are directed where they can have the greatest impact on the program and on the participants (Conger & Benjamin, 1999). It helps to develop training objectives that are tailored directly to address the obstacles and dilemmas impacting the implementation of the organization's strategic goals. Other types of assessment can work as well to enhance training. For example, a leadership effectiveness survey that is sent to the leadership team prior to training can provide information for the trainer on how the leadership team works together, communicates, or involve others in planning and decision making. The important point is that without pre-assessment measures, trainers will not totally understand the fundamental leadership challenges of the organization and may possibly develop a training program that is not what the leaders need most.

Conger & Benjamin (1999) recognized that managerial leadership development is "no longer focused on the individual learner but increasingly on shaping the worldviews and behaviors of cohorts of managers and, ... transforming even entire organizations" (p.

xii). Therefore, it is believed that the most effective leadership development programs are those that focus participants on the organization's strategic agenda and improve employees' understanding of how the agenda can be implemented. Four studies in this meta-analysis focused on strategic stewardship. However, one caveat is that for this type of training program to work, the strategic agenda must be clearly articulated at the organizational level. By enhancing participants' understanding of the organization's goals, it increases their capabilities, develops a shared organizational vision, and makes possible a wide-scale organizational change. According to Fiedler (1996), "we know very little about the processes of leadership and managerial training that contribute to organizational performance" (p. 244). Therefore, more training needs to occur that focuses on the strategic agenda of the organization.

It was no surprise that so few training programs focus on the top management level of organizations. For instance, in this meta-analysis only 13 out of 83 organizations tailored formal training programs for top managers, with the majority of those being in education (for principals). This research showed that managerial leadership development for the top levels of the organization was effective, and particularly those programs with behavior or system outcomes. There is a continued need for leadership development programs structured for top managers around specific learning objectives. It is so critical that senior management build a shared vision at the top that permeates the entire organization. Developing and implementing a shared vision is difficult especially in large multinational organizations where top leaders are located around the globe or must move to new positions frequently. This is obvious as only four studies in this research involved multinational companies. Nevertheless, critical strategic thinking skills, an understanding

of their own assumptions and those of their employees, and a greater insight into the organization as a whole should be important elements of leadership development programs for top managers.

Managerial leadership development programs should be designed to bring individuals from various levels and various units of the organization together to share their experiences and interpret information about the organization. Almost one half of the interventions in this meta-analysis (41) were with participants from mixed levels of responsibility, and those interventions were effective. Developing training programs using the high-performance leadership competency framework (Holton & Naquin, 2001) would integrate multiple leadership perspectives from all levels of the organization and connect organizational performance and leadership development. The multilevel approach would ensure that all levels of the organization have a consistent understanding of the strategic direction and the steps necessary to move the organization forward. The competency model allows for emphasis on different facets of leadership for individuals at different levels of management.

The management development literature indicates that training programs are commonly designed using multiple training techniques and multiple outcome measures (Collins, 2002; Conger & Benjamin, 1999). For instance, approximately one third of the studies in this meta-analysis had a combination of knowledge and expertise outcomes, and 82% measured system outcomes in conjunction with outcomes at the expertise or knowledge level.

It is important that organizations provide training programs with an organizational focus (both system and financial), focusing on strategic stewardship, visioning, and

transformational leadership. But, equally as important, organizations must spend time evaluating the effectiveness of those interventions with system outcomes and report the findings so that other organizations learn from them on what is or what is not effective. This would entail HRD professionals taking the lead to develop evaluation instruments that effectively measure performance level outcomes.

There has been a trend in the last 20 years toward using multiple evaluation techniques that include evaluations by subordinates, peers, supervisors and self-assessments). In this meta-analysis, 27 studies (33%) used a combination of both self and other measurements. In addition, some organizations use qualitative evaluation methods as well as quantitative methods. Multiple evaluation methods serve organizations well, and most specifically in wide-scale organization changes.

What is often overlooked regarding training but must be considered is the cost to the organization of trainees in the classroom – the return on investment made by the training program. This is important as large sums of money are invested in managerial leadership development programs annually (Gibler, Carter, & Goldsmith, 2000). The cost for higher paid managers to be in a classroom, away from work to attend the training is substantial. While it is known that training programs are effective, organizations should do a cost analysis to determine the actual return on investment from training initiatives.

### **Implications for Future Research**

The literature indicates that human resource development professionals are making efforts to enhance training effectiveness by conducting training needs assessments and creating the appropriate climate for transfer of training. In addition, they are conducting training evaluations beyond the reaction and learning levels to

demonstrate that training impacts organizational performance as well as performance of the individual employee. There has been a resurgence of interest in training evaluation (Alliger et al., 1997; Dionne, 1996; Holton, 1996; Moller & Mallin, 1996), with researchers exploring cause-effect relationships between interventions and the participants' learning, job performance, and system-level results. However, this meta-analysis pointed to the following as future research opportunities:

- 1) Anecdotal and case study evidence suggests that HRD should continue to provide job assignment experiences and developmental relationships that are linked to the organizational strategy as a preferred management development experience. However, more empirical studies are needed on the outcomes of feedback, on-the-job experiences, coaching and mentoring, as these interventions are the leading edge of managerial leadership development.
- 2) The literature search uncovered an emerging trend of transformational leadership research. However, there were surprisingly few research studies on the training of transformational leadership skills. The nature and application of transformational leadership, including how it can be developed and used to create cultures in which both people and performance are valued, cannot be measured by questionnaires alone. So, not only more training is needed in this area, but also HRD must develop qualitative as well as quantitative evaluation methods to produce the best results for assessing changes as a result of strategic development or transformational leadership interventions, and report those results.
- 3) Visioning is a topic that requires far greater empirical attention as a managerial leadership development intervention. More empirical work is needed in order to

develop a better understanding of potential antecedents of vision creation and to determine whether vision training actually works.

- 4) There is encouraging initial evidence of the effectiveness of team training (Eden, 1986; Graen, Novak, & Sommerkamp, 1982). However, given the widespread use of teams across organizations, more research is needed to provide definitive use to practitioners regarding the effectiveness of team training. HRD professionals should take the lead in designing techniques for learning team leadership skills, conduct further empirical research on team leadership, and report the results of those programs.
- 5) There is surprising little reported systematic evaluation of training programs with organizational performance as an outcome (Collins, 2001; Sogunro, 1997). While the prevailing principles of most management development literature are rooted in organizational strategy and organizational structure, the relationship between corporate performance and individual leadership lacks significant empirical support. For results to occur, the intervention must be linked with organizational goals and have utility or payoff to the organization (Swanson & Holton, 1999). HRD should take the lead by strategically aligning training and development systems that advance and sustain the organization's competitive position in its market.
- 6) Evaluation methods as known today are not sufficient for organizations to measure organizational-level performance improvement. Because organizations are facing a more competitive global economy with increased performance demands, HRD professionals should take the lead in combining evaluation theory

with performance-based management development theory to create the appropriate system for measurement of organizational-level performance. Evaluation methods must be specific, but yet broad enough, to satisfy the evaluation needs of all organizations while providing methods to conduct empirical research on outcomes of managerial leadership development programs.

- 7) The findings in this meta-analysis pointed to several variables as possible moderators of the effectiveness of managerial leadership development programs. However, because the number of studies was small in this meta-analysis, low power perhaps caused important moderators to remain undetected. Further analysis should be done in relation to research design, organization type, job classification level, and measurement methods as possible moderators of managerial leadership development interventions. Organizations should be encouraged to conduct more research and report findings so that an adequate data set is available to determine if there are moderators that may a difference in the effectiveness of programs.
- 8) More findings should be reported from pretest-posttest research design studies as this design is the only one that incorporates treatment by subject interaction, or the individual learner differences, in response to training. In addition, evaluations of behavioral change require a systematic assessment of job performance both before and after completion of the intervention. But, researchers typically use only control group design as they believe that other designs are flawed by threats to internal and external validity. However, often a SGPP design is the only type to measure training effectiveness.



- 9) Further analysis of PPWC studies should be conducted, as they provide an opportunity for further research by splitting the data set into POWC and SGPP. Carlson and Schmidt (1999) believe that PPWC is an outstanding research design as it captures more information than other designs. Because some meta-analyses exclude these studies, it appears that empirical research does not maximize the use of the additional data in this design.
- 10) Because previous meta-analyses did not provide an effect size based on an aggregation of studies with expertise-objective outcomes, the only known findings are based upon this research. Future research should separate behavioral outcomes that are measured subjectively from those measured objectively. In addition, future research should also separate results of studies with system outcomes from those with financial outcomes.

Meta-analysis is an evolving research strategy with refinements and technical improvements frequently being offered. However, there remains considerable controversy in meta-analysis methods for including multiple research designs, calculating effect size statistics, and conducting moderator detection and estimation. HRD researchers are strongly encouraged to use meta-analysis more often as a research technique, as findings from meta-analytic research has great potential for providing a basis of theory, guidelines, practices, and policies. Many researchers are unfamiliar with meta-analysis procedures. But, meta-analysis can regularly offer different findings and conclusions than conventional reviews of the same literature (Durlak and Lipsey, 1991).

Some suggestions for meta-analysis in managerial leadership research are:

- 1) Where treatment by subject interaction exists, further application of meta-analysis should be used to examine the effectiveness of interventions that use SGPP research design versus other designs.
- 2) Individuals who plan, implement, and report results of managerial leadership development studies in the future should:
  - a) Provide information regarding sample size, means, and standard deviations of comparison and experimental groups.
  - b) Report the type of test, alpha level, whether one-tailed or two-tailed, and degrees of freedom for each statistical test.
  - c) Provide detailed descriptive information about the characteristics of the intervention and the participants.
  - d) Provide numerical data when presenting graphs of statistical information.
  - e) Provide reliability data on measurement instruments.
  - f) Consider ways that both quantitative and qualitative methods can be combined, and provide statistical analyses of qualitative data.
  - g) Provide information regarding the organization type, training content, or participants that can enable more analyses of interventions. Report the magnitude of the treatment effect.

### **Summary**

In the past 20 years, the effectiveness of managerial leadership development programs has varied widely – some programs have been tremendously effective and

others have failed miserably. However, this research shows that organizations can feel comfortable that their managerial leadership development programs will produce substantial results, especially if they do the right development programs for the right people at the right time.

The most effective program for the organization is the one where the strategic framework of the organization drives the content of the training program. Serious problems arise when the goals and objectives of the training program are not linked to the overall strategy of the organization. To be effective, the training program must first be preceded by a needs assessment to ensure that the program objectives meet the specific needs of the organization and are linked to the organizational strategy. The most effective programs incorporate adult learning principles to enhance strategic thinking and other critical and complex learning capabilities. Programs must account for the individual learning styles of a diverse group of leaders and managers. And, leadership development is a never-ending process that involves multiple levels of the organization. Ultimately, the best approach to managerial leadership development is to focus on the organization's leadership requirements in the decade ahead rather than simply on those of today.

## REFERENCES

- Aguinis, H., & Pierce, C. A. (1998). Testing moderator variable hypotheses meta-analytically. Journal of Management, 24(5), 577-592.
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. Personnel Psychology, 42, 331-342.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. Personnel Psychology, 50(2), 341-359.
- Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. Organizational Dynamics, 25(2), 24-38.
- Avolio, B. J. (1999). Full leadership development: Building the vital forces in organizations. Thousand Oaks, CA: Sage.
- Avolio, B. J., & Howell, J. M. (1992). The impact of leadership behavior and leader-follower personality match on satisfaction and unit performance. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 225-236). Greensboro, NC: Center for Creative Leadership.
- Baker, D. E., Walsh, M. B., & Marjerison, L. (2000). Developing high performance leadership at the process level. In E. F. Holton & S. S. Naquin (Eds.), Developing high-performance leadership competency (Vol. 6, pp. 47-72). Baton Rouge, LA: Academy of Human Resource Development.
- Barling, J., Weber, T., & Kelloway, E. K. (1996). Effects of transformational leadership training on attitudinal and financial outcomes: A field experiment. Journal of Applied Psychology, 81(6), 827-832.
- Bass, B. M. (1985). Leadership and performance beyond expectations. New York: McGraw-Hill.
- Bass, B. M. (1990). Bass & Stodgill's handbook of leadership: Theory, research, and managerial application. New York: Free Press.
- Bass, B. M. (1998). Transformational leadership: Industry, military, and educational impact. Mahwah, NJ: Lawrence Erlbaum.
- Bassi, L. J., Benson, G., & Cheney, S. (1996). The top ten trends. Training and Development, 50, 28-42.
- Bates, R. A. (1999). Measuring performance improvement. In R. J. Torrance (Ed.), Performance improvement theory and practice. San Francisco: Berrett-Koehler.

Bayley, E. W. (1988). A meta-analysis of evaluations of the effect of continuing education on clinical practice in the health professions. Unpublished doctoral dissertation, University of Pennsylvania.

Beer, M. (1980). Organization change and development: A systems view. Santa Monica, CA: Goodyear Publishing Company.

Bennis, W., & Nanus, B. (1985). Leaders: The strategies for taking charge. San Francisco: Harper & Rowe.

Berkeley, R. (1988). Question: Do management development meetings change attitudes and behavior? Journal of Successful Meetings, 36 (4), 32-35.

Berry, J. K. (1990). Linking management development to business strategies. Training and Development, August 1990, 20-21.

Black, S., Morrison, A., & Gregersen, H. (1999). Global explorers: The next generation of leaders. New York: Routledge.

Blake, R. R., & Mouton, J. S. (1964). The managerial grid. Houston, TX: Gulf Publishing.

Borenstein, M., & Rothstein, H. (1999). Comprehensive meta-analysis: A computer program for research synthesis. Englewood, NJ: Biostat, Inc.

Boyett, J., & Boyett, J. (1998). The guru guide: The best ideas of the top management thinkers. New York: Wiley.

Bracken, D. W. (1994). Straight talk about multi-rater feedback. Training and Development, 28(9), 44-51.

Brake, T. (1997). The global leader: Critical factors for creating the world class organization. Chicago: Irwin Professional Publishing.

Brinkerhoff, R. O., & Montesino, M. U. (1995). Partnerships for training transfer: Lessons from a corporate study. Human Resource Development Quarterly, 6(3), 263-274.

Briscoe, D. R. (1995). International human resource management. Englewood Cliffs, NJ: Prentice Hall.

Brungardt, C. (1996). The making of leaders: A review of the research in leadership development and education. The Journal of Leadership Studies, 3(3), 81-95.

Brutus, S., Ruderman, M. N., Ohlott, P. J., & McCauley, C. D. (2000). Developing from job experiences: The role of organization-based self-esteem. Human Resource Development Quarterly, 11(4), 367-380.

Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. Journal of Applied Psychology, 71, 232-245.

Burke, W., & Litwin, G. (1992). A causal model of organizational performance and change. Journal of Management, 18(3), 523-545.

Burns, J. M. (1978). Leadership. New York: Harper-Rowe.

Caligiuri, P. M., & Stroh, L. K. (1995). Multinational corporation management strategies and international human resources practices: Bringing IHRM to the bottom line. The International Journal of Human Resource Management, 6(3), 494-507.

Campbell, J. A., Dunnette, M. D., Lawler, E. E., & Weick, K. E. (1970). Managerial behavior, performance, and effectiveness. New York: McGraw Hill.

Carlisle, K. E., & Henrie, D. (1993). Are you doing high-impact HR? Training and Development, August 1993, 47-53.

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. Journal of Applied Psychology, 84(6), 851-862.

Carnevale, A. P., & Schulz, E. R. (1990, July). Return on investment: Accounting for training. Training and Development Journal, S2-S32.

Cato, B. (1990). Effectiveness of a management training institute. Journal of Park and Recreation Administration, 8(3), 38-46.

Chen, L. (1994). Cross-cultural training effectiveness: A meta-analysis. Unpublished masters thesis, Central Michigan University.

Church, A. H., & Bracken, D. W. (1997). Advancing the state of the art of 360-degree feedback. Journal of Applied Psychology, 27(11), 983-1020.

Clark, K. C., Clark, M. B., & Campbell, D. P. (Eds.) (1992). Impact of leadership. Greensboro, NC: Center for Creative Leadership.

Clark, K. C., & Clark, M. B. (1994). Choosing to lead. (2<sup>nd</sup> Edition). Greensboro, NC: Center for Creative Leadership.

Clement, R. W. (1982). Testing the hierarchy theory of training evaluation: An expanded role for trainee reactions. Public Personnel Management Journal, 11, 176-184.

Collins, D. B. (2001). Organizational performance: The future focus of leadership development programs. The Journal of Leadership Studies, 7(4), 43-54.

Collins, D. B. (2002). Performance-level evaluation methods in management development from 1986 – 2000. Human Resource Development Review, 1(1), 91-110.

Collins, D. B., Lowe, J. S., & Arnett, C. R. (2000). In E. F. Holton & S. S. Naquin (Eds.), Developing high-performance leadership competency (Vol. 6, pp. 18-46). Baton Rouge, LA: Academy of Human Resource Development.

Conger, J. A. (1992). Learning to lead: The art of transforming managers into leaders. San Francisco: Jossey-Bass.

Conger, J. A., & Benjamin, B. (1999). Building leaders. San Francisco: Jossey-Bass Publishers.

Cooper, H. M. (1984). The integrative research review: A systematic approach. Beverly Hills: Sage.

Daugherty, R. A., & Williams, S. E. (1997). The long-term impact of leadership development: An assessment of a statewide program. The Journal of Leadership Studies, 4(2), 101-115.

Dionne, P. (1996). The evaluation of training activities: A complex issue involving different stakes. Human Resource Development Quarterly, 7(3), 279-286.

Dixon, N. M. (1990). The relationship between trainee responses on participation reaction forms and posttest scores. Human Resource Development Quarterly, 1, 129-137.

Dowling, P. J., Welch, D. E., & Schuler, R. S. (1999). International human resource management: Managing people in a multinational context. Cincinnati: South-Western College Publishing.

Downton, J. V. (1973). Rebel leadership: Commitment and charisma in a revolutionary process. New York: Free Press.

Driggs, G. E. (1999). Outcomes training goes formal. Behavioral Health Management, 19(4), 32-33.

Durlak, J. A., & Lipsey, M. W. (1991). A practitioner's guide to meta-analysis. American Journal of Community Psychology, 19(3), 291-332.

Evans, M. G. (1970). The effects of supervisory behavior on the path-goal relationship. Organizational Behavior and Human Performance, 5, 277-298.

Facteau, C. L., Facteau, J. D., Schoel, L. C., Russell, J. E. A., & Poteet, M. L. (1998). Reactions of leaders to 360-degree feedback from subordinates and peers. Leadership Quarterly, 9(4), 427-448.

Faerman, S. R., & Ban, C. (1993). Trainee satisfaction and trainee impact: Issues in training evaluation. Public Productivity and Management Review, 16(3), 299-314.

Fiedler, F. E. (1964). A contingency model of leadership effectiveness. In L. Berkowitz (Ed.), Advances in experimental social psychology, Vol. 1 (pp. 149-190). New York: Academic Press.

Fiedler, F. E. (1992). The role and meaning of leadership. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 95-106). Greensboro, NC: Center for Creative Leadership.

Fiedler, F. E. (1996). Research on leadership selection and training: One view of the future. Administrative Science Quarterly, 41, 241-250.

Fiedler, F. E., & Chemers (1984). Leadership effectiveness: The leader match concept (2<sup>nd</sup> ed.). New York: John Wiley.

Fiedler, F.E., Chemers, M.M. and Mahar, L. (1976). Improving Leadership Effectiveness: The Leader Match Concept, New York: John Wiley and Sons.

Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. Leadership Quarterly, 2, 245-287.

Friedman, T. L. (2000). The lexis and the olive tree. New York: Anchor Books.

Fulmer, R. M., & Wagner, S. (1999). Leadership: Lessons from the best. Training and Development, 53(3), 28-33.

Gardner, J. W. (1990). On leadership. New York: Free Press.

Gibler, D., Carter, L., & Goldsmith, M. (2000). Best practices in leadership development handbook. San Francisco: Jossey-Bass.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. Educational Researcher, 5, 3-8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Goldstein, I. L. (1980). Training in work organizations. Annual Review of Psychology, 31, 229-270.



Goldstein, I. L. (1986). Training in organizations: Needs assessment, design, and evaluation. Monterrey, CA: Brooks-Cole.

Gordon, J. (1987, June). Romancing the bottom line. Training, 31-42.

Graen, G. (1976). Role making processes within complex organizations. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago, IL: Rand McNally.

Guns, B. (1996). The faster learning organization: Gain and sustain the competitive edge. San Diego, CA: Pfeiffer.

Hackman, J. R., & Walton, R. E. (1986). Leading groups in organizations. In P. S. Goodman & Associates (Eds.), Designing effective work groups (pp. 72-119). San Francisco: Jossey-Bass.

Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. Human Resource Management, 32(2 & 3), 325-351.

Herling, R. W. (2000). Operational definitions of expertise and competence. In R. W. Herling & J. Provo (Eds.), Strategic perspectives of knowledge, competence, and expertise (pp. 8-21). San Francisco: Berrett-Koehler.

Hersey, P., & Blanchard, K. H. (1969). Life-cycle theory of leadership. Training and Development Journal, 23, 26-34.

Hofstede, G. (1986). Cultural differences in teaching and learning. International Journal of Intercultural Relations, 10, p. 301-320.

Holton, E. F., III (1996). The flawed four-level evaluation model. Human Resource Development Quarterly, 7(1), 5-21.

Holton, E. F. (1999). Performance domains and their boundaries. In R. J. Torraco (Ed.), Advances in Developing Human Resources, Vol. 1 (pp. 26-46). Baton Rouge, LA: Academy of Human Resource Development.

Holton, E. F., & Naquin, S. S. (2000). Developing high performance leadership competency. Baton Rouge, LA: Academy of Human Resource Development.

Hooijberg, R., Hunt, J. G., & Dodge, G. E. (1997). Leadership complexity and development of the leaderplex model. Journal of Management, 23, 375-408.

House, R. J. (1971). A path-goal theory of leader effectiveness. Administrative Science Quarterly, 16, 321-328.

House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership. Quo vadis? Journal of Management, 23, 409-473.

Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee job performance and work force productivity. American Psychologist, 38, 473-478.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis. Newbury Park, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage Publications.

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. Academy of Management Journal, 38(3), 635-672.

Ireland, R., & Hitt, M. (1999). Achieving and maintaining strategic competitiveness in the 21<sup>st</sup> century: The role of strategic leadership. The Academy of Management Executives, 13(1), 43-57.

Jacobs, R. L., & Jones, M. J. (1995). Structures on-the-job training: Unleashing employee expertise in the workplace. San Francisco: Berrett-Koehler.

Johnson, R. R., McLaughlin, S. D., Saari, L. M., & Zimmerle, D. M. (1988). The demand for university-based executive education. The Magazine of the Graduate Admissions Council, 4(3), 11-29.

Kanter, R. (1997). Restoring people to the heart of the organization of the future. In F. Hesselbein, M. Golesmith, & R. Beckhard (Eds.). The Organization of the future. San Francisco: Jossey-Bass.

Kaplan, R. S., & Norton, D. P. (1993). Putting the balanced scorecard to work. Harvard Business Review, September-October 1993, 134-146.

Katz, D., & Kahn, R. (1978). The social psychology of organizations. New York: John Wiley & Sons.

Kaufman, R., & Watkins, R. (1996). Cost-consequences analysis. Human Resource Development Quarterly, 7(1), 87-100.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), Training and Development Handbook (2<sup>nd</sup> Ed., Chapter 18). New York: McGraw-Hill.

Kirkpatrick, D. L. (1996). Invited reaction: Reaction to Holton article. Human Resources Development Quarterly, 7, 5-15.

Kirkpatrick, D. L. (1998). Evaluating training programs: The four levels. San Francisco: Berrett-Koehler.

Kirkpatrick, S. A., & Locke, E. A. (1991). Leadership: Do traits matter? The Executive, 5, 48-60.

Klenke, K. (1993). Leadership education at the great divide: Crossing into the 21<sup>st</sup> century. Journal of Leadership Studies, 1(1), 112-127.

Kline, P., & Saunders, B. (1993). 10 steps to a learning organization. Arlington, VA: Great Ocean Publishers.

Knauff, E. B. (1992). Leadership, organizational culture, and organizational outcomes. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 37-46). Greensboro, NC: Center for Creative Leadership.

Koch, M. J., & McGrath, R. G. (1996). Improving labor productivity: Human resource policies do matter. Strategic Management Journal, 17(5), 335-354.

Kotter, J. P. (1990). A forceful change: How leadership differs from management. New York: Free Press.

Kotter, J. (1996). What leaders really do. In R. Steers, L. Porter, & G. Bigley (Eds.), Motivation and leadership at work. New York: McGraw-Hill.

Kouzes, J. M., & Posner, B. Z. (1996). The leadership challenge: How to get extraordinary things done in organizations. San Francisco: Jossey-Bass.

Kraiger, K., Ford, J., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. Journal of Applied Psychology, 75(2), 311-328.

Krohn, R. A. (2000). Training as a strategic investment. In R. W. Herling & J. Provo (Eds.), Strategic perspectives on knowledge, competence, and expertise. San Francisco: Berrett-Koehler.

Kuchinke, P. (1995). Managing learning for performance. Human Resource Development Quarterly, 6(3), 307-316.

Kuchinke, K. P. (2000). The role of feedback in management training settings. Human Resource Development Quarterly, 11(4), 381-401.

Lai, L. C. (1996). A meta-analysis of research on the effectiveness of leadership training programs. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Lau, R. S., & May, B. E. (1998). A win-win paradigm for quality of work life and business performance. Human Resource Development Quarterly, 9(3), 211-225.

Lam, L. L., & White, L. P. (1998). Human resource orientation and corporate performance. Human Resource Development Quarterly, 9(4), 351-364.

Larson, C. E., & LaFasto, F. M. J. (1989). Teamwork: What must go right/What can go wrong. Newbury Park, CA: Sage.

Leddick, A. S. (1987). Effects of training on measures of productivity: A meta-analysis of the findings of forty-eight experiments. Unpublished doctoral dissertation, Western Michigan University.

Lepsinger, R., & Lucia, A. D. (1997). The art and science of 360-degree feedback. San Francisco: Pfeiffer.

Lercel, M. E., & Field, L. M. (1998). Preparing for the transfer of leadership: Into the void or into a well-planned future? Canadian Manager, 24(1), 23-27.

Levi, A. S., & Mainstone, L. E. (1992). Breadth, focus, and content in leader priority-setting: Effects on decision quality and perceived leader performance. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 429-442). Greensboro, NC: Center for Creative Leadership.

Lipsey, M. S., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. American Psychologist, 48, 1181-1209.

Lipsey, M. W. (1990). Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage.

Lohmann, D. (1992). The impact of leadership on corporate success: A comparative analysis of the American and Japanese experience. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 59-80). Greensboro, NC: Center for Creative Leadership.

Lynham, S. A. (2000). Leadership Development: A review of the theory and literature. In P. Kuchinke (Ed.), Proceedings of the 2000 Academy of Human Resource Development Annual Meeting. Baton Rouge, LA: Academy of Human Resource Development.

Lynham, S. A., & Swanson, R. A. (1997). The development and evaluation of a model of responsible leadership for performance: Beginning the journey. In Proceedings of the Academy of Human Resource Development (pp. 22-23). Baton Rouge, LA: Academy of Human Resource Development.

Marquardt, M. J., & Engel, D. W. (1993). Global human resource development. Upper Saddle River, NJ: Prentice Hall.

Marson, P. P., & Bruff, C. D. (1992). The impact of classroom leadership training on managerial/supervisory job performance. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 359-364). Greensboro, NC: Center for Creative Leadership.

Martin, L. L., & Kettner, P. M. (1997). Performance measurement: The new accountability. Administration in Social Work, 21(1), 17-29.

Maznevski, M. L., & DiStefano, J. J. (2000). Global leaders are team players: Developing global leaders through membership on global teams. Human Resource Planning, 39(2, 3), 195-208.

McCall, M. W., Jr. (1998). High flyers: Developing the next generation of leaders. Boston: Harvard Business School Press.

McCall, M. W., Jr., Lombardo, M. M., & Morrison, A. M. (1988). The lessons of experience: How successful executives develop on the job. New York: Lexington Books.

McCauley, C. D., & Brutus, S. (1998). Management development through job experiences. Greensboro, NC: Center for Creative Leadership.

McCauley, C. D., Moxley, R. S., & Van Velsor, E. (1998). The Center for Creative Leadership handbook of leadership development. San Francisco: Jossey-Bass.

McCauley, C. D., Ruderman, M. N., Ohlott, P. J., & Morrow, J. E. (1994). Assessing the developmental components of managerial jobs. Journal of Applied Psychology, 79, 544-560.

McLagan, P. A. (1997). Competencies: The next generation. Training and Development, 51, 40-47.

Meindl, J. R., & Ehrlich, S. B. (1987). The romance of leadership and the evaluation of organizational performance. Academy of Management Journal, 30, 91-109.

Meisel, S., & Fearon, D. (1996). Leading learning. In S. Cavaleri & D. Fearon (Eds.), Managing in organizations that learn. Cambridge, MA: Blackwell.

Meldrum, M., & Atkinson, S. (1998). Is management development fulfilling its organizational role? Management Decision, 36(7-8), 528-533.

Mitchell, T. R. (1983). The effects of social, task, and situational factors on motivation, performance, and appraisal. In F. J. Landry, S. Zedeck, & J. Cleveland (Eds.). Performance measurement and theory (pp. 29-59). Hillsdale, NJ: Erlbaum.

Moller, L., & Mallin, P. (1996). Evaluation practices of instructional designers and organizational supports and barriers. Performance Improvement Quarterly, 9(4), 82-92.

Moxnes, P., & Eilertsen, D. (1991). The influence of management training upon organizational climate: An exploratory study. Journal of Organizational Behavior, 12, 399-411.

Newstrom, J. W. (1995). Review of "Evaluating training programs: The four levels of D. L. Kirkpatrick." Human Resource Development Quarterly, 6, 417-319.

Northouse, P. G. (1997). Leadership: Theory and practice. Thousand Oaks, CA: Sage.

Olian, J. D., Durham, C. C., Kristoff, A. L., Brown, K. G., Pierce, R. M., & Kunder, L. (1998). Designing management training and development for competitive advantage: Lessons from the best. Human Resource Planning, 21(1), 20-33.

Osman-Gani, A. M. (2000). Developing expatriates for the Asia-Pacific region: A comparative analysis of multinational enterprise managers from five countries across three continents. Human Resource Development Quarterly, 11(3), 213-235.

Penwell, L. W. (1992). The impact of leaders on group and intergroup dynamics in teleconferences between interdependent work teams. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 477-486). Greensboro, NC: Center for Creative Leadership.

Pfeiffer, J. (1994). Competitive advantage through people: Unleashing the power of the work force. Boston: Harvard Business School Press.

Phillips, J. J. (1997). Handbook of training evaluation and measurement methods. Houston, TX: Gulf.

Plant, R. A., & Ryan, R. J. (1992). Training evaluation: A procedure for validating an organization's investment in training. Journal of European Industrial Training, 16(10), 22-38.

Porter, M. E. (1985). The competitive advantage: Creating and sustaining performance. New York: Free Press.

Preskill, H. J. (1997). HRD evaluation as the catalyst for organizational learning. In E. F. Holton, III (Ed.), Proceedings of the Academy of Human Resource Development (pp. 49-56). Baton Rouge, LA.

Provo, J. (2000). Measuring human capital. In R. W. Herling & J. Provo (Eds.), Strategic perspectives on knowledge, competence, and expertise. San Francisco: Berrett-Koehler.

Quast, L. N., & Hazucha, J. F. (1992). The relationship between leaders' management skills and their groups' effectiveness. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 199-214). Greensboro, NC: Center for Creative Leadership.

Reichheld, F. F. (1996). The loyalty effect: The hidden force behind growth, profits, and lasting value. Boston: Harvard Business School Press.

Riechmann, D. (1992). High-involvement, high-performance teams in higher education: The impact of leadership. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 257-268). Greensboro, NC: Center for Creative Leadership.

Rolls, J. (1995). The transformational leader: The wellspring of the learning organization. In S. Chawla & J. Rensch (Eds.), Creating the learning organization. Alexandria, VA: American Society of Training and Development.

Rosenbaum, J., & Keller, R. A. (1995). Building a learning organization at Coopers and Lybrand. Planning Review, 22, 28-29.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. Psychological Bulletin, 86, 85-97.

Rosenthal, R. (1984). Meta-analysis procedures for social research. Beverly Hills, CA: Sage.

Rost, J. C. (1991). Leadership for the twenty-first century. New York: Praeger.

Rost, J. C. (1993). Leadership for the twenty-first century. Westport, CT: Greenwood.

Rummler, G. A., & Brache, A. P. (1995). Improving performance: How to manage the white space on the organizational chart. San Francisco: Jossey-Bass.

Ruona, W. E. A., Leimbach, M., Holton, E. F., III (1999). The relationship between learner utility reactions and predictors of learning transfer: Implications for evaluation. Academy of Human Resource Development. (p. 1100-1106).

Saari, L. M., Johnson, T. R., McLaughlin, S. D., & Zimmerle, D. M. (1988). A survey of management training and education practices in U.S. companies. Personnel Psychology, 41, 731-743.

Sackett, P. R. (in press). The status of validity generalization research: Key issues in drawing inferences from cumulative research findings. In K. R. Murphy (Ed.), Validity generalization: A critical review. Mahwah, NJ: Erlbaum.

Senge, P. (1990). The fifth discipline: The art and science of the learning organization. New York: Doubleday.

Senge, P. (1992, March). Building the learning organization. Journal for Quality and Participation, 30-38.

Smith, M. (1987). Sizing up your training. Journal of Personnel Administrator, 32(4), 62-65.

Sogunro, O. A. (1997). Impact of training on leadership development: Lessons from a leadership training program. Evaluation Review, 21(6), 713-737.

Sorcher, M., & Goldstein, A. P. (1972). A behavioral modeling approach in training. Personnel Administration, 35, 35-41.

Sourcie, D. (1994). Effective managerial leadership in sport organizations. Journal of Sport Management, 8, 1-19.

Spreitzer, G. M., McCall, M. W., & Mahoney, J. D. (1997). Early identification of international executive potential. Journal of Applied Psychology, 82(1), 6-29.

Svyantek, D. J., & DeShon, R. P. (1992). Leaders and organizational outcomes in established industries: An analysis of Lee Iacocca and the American automobile industry. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 293-304). Greensboro, NC: Center for Creative Leadership.

Swanson, R. A. (1994). Analysis for improving performance: Tools for diagnosing organizations and documenting workplace expertise. San Francisco: Berrett-Koehler.

Swanson, R. A. (1998). Demonstrating the financial benefit of human resource development: Status and update on the theory and practice. Human Resource Development Quarterly, 9(3), 285-295.

Swanson, R. A., & Gradous, D. B. (1988). Forecasting financial benefits of human resource development. San Francisco: Jossey-Bass.



Swanson, R. A., & Holton, E. F. (1999). Results: How to assess performance, learning, and perceptions in organizations. San Francisco: Berrett-Koehler.

Swanson, R. A., & Mattson, B. W. (1997). Development and validation of the critical outcome technique (COT). In R. Torraco (Ed.), Academy of Human Resource Development 1996 Annual Proceedings.

Tannenbaum, S. I. (1997). Enhancing continuous learning: Diagnostic findings from multiple companies. Human Resource Management, 36, 437-52.

Teal, T. (1998). The human side of management. In Harvard Business Review on Leadership. Boston: Harvard Business School.

Thoms, P., & Klein, H. J. (1994). Participation and evaluative outcomes in management training. Human Resource Development Quarterly, 5(1), 27-39.

Tichy, N., & Devanna, M. A. (1990). The transformational leader. New York: John Wiley & Sons.

Torraco, R. J. (1999). Performance improvement theory and practice. San Francisco: Berrett-Koehler.

Torraco, R. J., & Swanson, R. A. (1995). The strategic roles of human resource development. Human Resource Planning, 18(4), 10-21.

Tucker, M. L., Bass, B. M., & Daniel, L. G. (1992). Transformational leadership's impact on higher education satisfaction, effectiveness, and extra effort. In K. E. Clark, M. B. Clark, & D. P. Campbell (Eds.), Impact of leadership (pp. 169-176). Greensboro, NC: Center for Creative Leadership.

Tung, R., & Miller, E. (1990). Managing in the twenty-first century: The need for global orientation. Management International Review, 5-18.

Ulrich, D. (1997). Measuring human resources: An overview of practice and a prescription for results. Human Resource Management, 36, 303-320.

Ulrich, D., Zenger, J., & Smallwood, N. J. (1999). Results-based leadership. Boston: Harvard Business School Press.

Vaill, P. B. (1990). Managing as a performing art: New ideas for a world of chaotic change. San Francisco: Jossey-Bass.

Waldman, D. A., Ramirez, G. G., House, R. J., & Puranam, P. (2001). Does leadership matter? CEO leadership attributes and profitability under conditions of perceived environmental uncertainty. Academy of Management Journal, 44(1), 134-143.

Walker, J. W., & Bechet, T. P. (1991). Defining effectiveness and efficiency measures in the context of human resource strategy. In R. J. Niehaus & K. F. Price (Eds.), Bottom line results from strategic resource planning (pp.235-260). New York: Plenum Press.

Wanous, J. P., Sullivan, S. E., & Malinak, J. (1989). The role of judgment calls in meta-analysis. Journal of Applied Psychology, 74(2), 259-264.

Wexley, K. N. (1984). Personnel training. Annual Review of Psychology, 35, 519-551.

Wofford, J. C., & Liska, L. Z. (1993). Path-goal theories of leadership: A meta-analysis. Journal of Management, 19(4), 857-876.

Wortman, P. M. (1994). Judging research quality. In H. Cooper and L. V. Hedges (Eds.). The handbook of research synthesis (pp. 97-109). New York: Russell Sage Foundation.

Yeaton, W. H., & Wortman, P. M. (1984). On the reliability of meta-analytic reviews: The role of intercoder agreement. Evaluation Review, 17, 292-309.

Yeung, A., & Ready, D. (1995). Developing leadership capabilities of global corporations: A comparative study in eight nations. Human Resource Management, 34(4), 529-547.

Yukl, G. (1989). Managerial leadership: A review of theory and research. Journal of Management, 15, 251-289.

Yukl, G. A. (1994). Leadership in organizations. Englewood Cliffs, NJ: Prentice Hall.

Yukl, G., & van Fleet, D. D. (1992). Theory and research on leadership in organizations. In M. D. Dunnett & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (Vol. 3, pp. 147-197). Palo Alto, CA: Consulting Psychologists Press.

Zhang, J. (1999). Effects of management training on trainees' learning, job performance, and organization results: A meta-analysis of evaluation studies from 1983-1997. Unpublished doctoral dissertation, Oklahoma State University.

## APPENDIXES

**APPENDIX A**  
**SEARCH FOR UNPUBLISHED STUDIES**

Dear \_\_\_\_\_,

I am conducting a meta-analysis on the effectiveness of managerial leadership development programs from 1982-2001 for my doctoral research.

I intend to include unpublished studies in my meta-analysis and I need your assistance in locating unpublished research articles. These may be ones that you personally have written, or ones written by a fellow colleague. A copy of the article you have written, or the names and addresses of whom I should contact would be most appreciated.

For purposes of my meta-analysis, managerial leadership development is defined as a full-range of leadership development interventions including, but not limited to, mentoring, coaching, dyadic relationships, leader-follower relationships, multi-rater feedback, on-the-job experiences, job assignments, and formal training programs. To be included in the study sample, individual studies must incorporate an intervention that involves individuals at the managerial level with titles such as managers, leaders, executives, principals, officers, and/or supervisors, and an intervention that is a deliberately planned effort by an individual, group, or organization that enhances the managerial leadership potential at the individual, group or team, or organizational level. The outcome(s) of each study must be identifiable and the study must have quantitative statistics from which to extrapolate an effect size.

I thank you in advance for your assistance and advice. You may communicate me at the addresses and phone numbers listed below.

Sincerely,

Doris B. Collins  
Associate Vice Chancellor  
Student Life and Academic Services  
134 Johnston Hall  
Louisiana State University  
Baton Rouge, LA 70803

Phone: 225-578-0481  
Fax: 225-578-5036  
E-mail: [dcollin@lsu.edu](mailto:dcollin@lsu.edu)

## APPENDIX B

### META-ANALYSIS SAMPLE

Alsamani-Abdulrahman, S. (1998). The management development program in Saudi Arabia and its effect on managerial job performance. Unpublished doctoral dissertation, Mississippi State University.

Bankston, J. R. (1993). Instructional leadership behaviors of a selected group of principals in northeast Texas. Unpublished doctoral dissertation, East Texas State University.

Barling, J., Weber, T., & Kelloway, E. K. (1996). Effects of transformational leadership training on attitudinal and financial outcomes: A field experiment. Journal of Applied Psychology, 81(6), 827-832.

Bendo, J. J. (1984). Application of counseling and communication techniques to supervisory training in industry. Unpublished doctoral dissertation, University of Akron.

Birkenbach, X. C., Karnfer, L., & ter Morshuizen, J. D. (1985). The development and the evaluation of a behaviour-modelling training programme for supervisors. South African Journal of Psychology, 15(1), 11-19.

Briddell, W. B. (1986). The effects of a time management training program upon occupational stress levels and the Type A behavioral pattern in college administrators. Unpublished doctoral dissertation, Florida State University.

Bruwelheide, L. R., & Duncan, P. K. (1985). A method for evaluating corporation training seminars. Journal of Organizational Behavior Management, 7(1-2), 65-94.

Cato, B. (1990). Effectiveness of a management training institute. Journal of Park and Recreation Administration, 8(3), 38-46.

Clark, B. R. (1989). Professional development emphasizing the socialization progress for academicians planning administrative leadership roles. Unpublished doctoral dissertation, George Washington University.

Clark, H. B., Wood, R., Kuehnel, T., Flanagan, S., Mosk, M., & Northrup, J. T. (1985). Preliminary validation and training of supervisory interactional skills. Journal of Organizational Behavior Management, 7(1-2), 95-115.

Colan, N., & Schneider, R. (1992). The effectiveness of supervisor training: One- year follow-up. Journal of Employee Assistance Research, 1(1), 83-95.

Couture, S. (1987). The effect of an introductory management inservice program on the knowledge of management skills of assistant head nurses at a university teaching

hospital. Unpublished master's thesis, The University of Texas Medical Branch at Galveston.

Davis, B. L., & Mount, M. K. (1984). Effectiveness of performance appraisal training using computer assisted instruction and behavior modeling. Personnel Psychology, 37, 439-452.

Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. Journal of Applied Psychology, 74(4), 580-590.

DeNisi, A. S., & Peters, L. H. (1996). Organization of information in memory and the performance appraisal process: Evidence from the field. Journal of Applied Psychology, 81(6), 717-737.

DePiano, L. G., & McClure, L. F. (1987). The training of principals for school advisory council leadership. Journal of Community Psychology, 15, 253-267.

Devlin-Scherer, W. L., Devlin-Scherer, R., Wright, W., Roger, A. M., & Meyers, K. (1997). The effects of collaborative teacher study groups and principal coaching on individual teacher change. Journal of Classroom Interaction, 32(1), 18-22.

Donohoe, T. L., Johnson, J. T., & Stevens, J. (1997). An analysis of an employee assistance supervisory training program. Employee Assistance Quarterly, 12(3), 25-34.

Dorfman, P. W., Stephan, W. G., & Loveland, J. (1986). Performance appraisal behaviors: Supervisor perceptions and subordinate reactions. Personnel Psychology, 39(3), 579-597.

Dvir, T., Eden, D., Avolio, B. J., & Shamir, B. (2001). Impact of transformational leadership on follower development and performance: A field experiment. Manuscript accepted for publication, Academy of Management Journal.

Earley, P. C. (1987). Intercultural training for managers: A comparison of documentary and interpersonal methods. Academy of Management Journal, 30, 685-698.

Eden, D. (1986). Team Development: Quasi-experimental confirmation among combat companies. Group and Organization Studies, 11(3), 133-146.

Eden, D., Geller, D., Gewirtz, A., Gordon-Terner, R., Inbar, I., Liberman, M., Pass, Y., Salomon-Segev, I., & Shalit, M. (2000). Implanting pygmalion leadership style through workshop training: Seven field experiments. Leadership Quarterly, 11(2), 171-210.

Edwards, W. S. (1992). Training secondary school administrators for crisis management. Unpublished doctoral dissertation, University of Georgia.

Faerman, S. R., & Ban, C. (1993). Trainee satisfaction and training impact: Issues in training evaluation. Public Productivity and Management Review, 16(3), 299-314.

Frost, Dean E. (1986). A test of situational engineering for training leaders. Psychological Reports, 59(2), 771-782.

Fuller, A. G. (1985). A study of an organizational development program for leadership effectiveness in higher education. Unpublished doctoral dissertation. Memphis State University.

Gerstein, L. H., L. H., Eichenhofer, D. J., Bayer, G. A., Valutis, W., & Janowski, J. (1989). EAP referral training and supervisors' beliefs about troubled workers. Employee Assistance Quarterly, 4(4), 15-30.

Graen, G., Novak, M., & Sommerkamp, P. (1982). The effects of leader-member exchange and job design on productivity and satisfaction: Testing a dual attachment model. Organizational Behavior and Human Performance, 30, 109-131.

Haccoun, R. R., & Hamtiaux, T. (1994). Optimizing knowledge tests for inferring learning acquisition levels in single group training evaluation designs: The internal referencing strategy. Personnel Psychology, 47, 593-604.

Harrison, J. K. (1992). Individual and combined effects of behavior modeling and the cultural assimilator in cross-cultural management training. Journal of Applied Psychology, 77(6), 952-962.

Henry, J. E. (1983). The effects of team building, upper management support, and level of manager on the application of management training in performance appraisal skills as perceived by subordinates and self. Unpublished doctoral dissertation, University of Colorado.

Hill, S. (1992). Leadership development training for principals: Its initial impact on school culture. Unpublished doctoral dissertation, Baylor University.

Innami, I. (1994). The quality of group decisions, group behavior, and intervention. Organizational Behavior and Human Decision Processes, 60, 409-430.

Ivancevich, J. M. (1982). Subordinates' reactions to performance appraisal interviews: A test of feedback and goal-setting techniques. Journal of Applied Psychology, 67(5), 581-587.

Jalbert, N. M., Morrow, C., & Joglekar, A. (2000). Measuring change with 360's: Using a degree of change measure versus performance. Paper presented at the Fifteenth Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, April 14-16, 2000.

Josefowitz, A. J. (1984). The effects of management development training on organizational climate. Unpublished doctoral dissertation, University of Minnesota.

Katzenmeyer, M. H. (1988). An evaluation of behavior modeling training designed to improve selected skills of educational managers. Unpublished doctoral dissertation, Florida State University.

Klein, E. B., Kossek, E. E., & Astrachan, J. H. (1992). Affective reactions to leadership education: An exploration of the same-gender effect. Journal of Applied Behavioral Science, 28(1), 102-117.

Lafferty, B. D. (1998). Investigation of a leadership development program: An empirical investigation of a leadership development program. Unpublished doctoral dissertation, George Washington University.

Larkin, J. M. (1996). Effects of a conflict management course on nurse managers. Unpublished master's thesis, Duquesne University.

Larsen, J. E. (1983). An evaluative study of the effect of self-management training on group leadership training program for hospital managers. Unpublished doctoral dissertation, University of Nebraska.

Lawrence, H. V., & Wiswell, A. K. (1993). Using the work group as a laboratory for learning: Increasing leadership and team effectiveness through feedback. Human Resource Development Quarterly, 4(2), 135-148.

Martineau, J. W. (1995). A contextual examination of the effectiveness of a supervisory skills training program. Unpublished doctoral dissertation, Pennsylvania State University.

Mattox, R. J. (1985). Self, subordinate, and supervisor reports on the long-term effects of a management training program. Unpublished doctoral dissertation, East Texas State University.

Maurer, S. D., & Fay, C. (1988). Effect of situational interviews, conventional structured interviews, and training on interview rating agreement: An experimental analysis. Personnel Psychology, 41(2), 329-344.

May, G. L., & Kahnweiler, W. M. (2000). The effect of a mastery practice design on learning and transfer in behavior modeling training. Personnel Psychology, 53(2), 353-373.

McCauley, C. D., & Hughes-James, M. W. (1994). An evaluation of the outcomes of a leadership development program. Greensboro, NC: Center for Creative Leadership.



McCredie, H., & Hool, J. (1992). The validation, learning dynamics and application of an off-the-shelf course in analytical skills. Management Education and Development, 23(4), 292-306.

Moxnes, P., & Eilertsen, D. (1991). The influence of management training upon organizational climate: An exploratory study. Journal of Organizational Behavior, 12, 399-411.

Nelson, E. P. (1990). The impact of the Springfield development program on principals' administrative behavior. Unpublished doctoral dissertation, University of Minnesota.

Niska, J. M. (1991). Examination of a cooperative learning supervision training and development model. Unpublished doctoral dissertation, Iowa State University.

Paquet, B., Kasl, E., Weinstein, L., & Waite, W. (1987, May). The bottom line. Training and Development Journal, 41(5), 27-33.

Parker, S. K. (1998). Enhancing role breadth self-efficacy: The roles of job enrichment and other organizational interventions. Journal of Applied Psychology, 83(6), 835-852.

Posner, C. S. (1982). An inferential analysis measuring the cost effectiveness of training outcomes of a supervisory development program using four selected economic indices: An experimental study in a public agency setting, Unpublished doctoral dissertation, Ball State University.

Reaves, W. M. (1993). The impact of leadership training on leadership effectiveness. Unpublished doctoral dissertation, North Carolina State University.

Robertson, M. M. (1992). Crew resource management training program for maintenance technical operators: Demonstration of a systematic training evaluation model. Unpublished doctoral dissertation, University of Southern California.

Rosti, R. T., Jr., & Shipper, F. (1998). A study of the impact of training in a management development program based on 360 feedback. Journal of Managerial Psychology, 13(1-2), 77-89.

Russell, J. S., Wexley, K. N., & Hunter, J. E. (1984). Questioning the effectiveness of behavior modeling training in an industrial setting. Personnel Psychology, 37, 465-481.

Savan, M. (1983). The effects of leadership training program on supervisory learning and performance. Unpublished doctoral dissertation, The Union for Experimenting Colleges and Universities, Cincinnati, Ohio.

Scandura, T. A., & Graen, G. B. (1984). Moderating effects of initial leader-member exchange status on the effects of a leadership intervention. Journal of Applied Psychology, 69(3), 428-436.

Schneider, R. J., & Colan, N. B. (1992). The effectiveness of employee assistance program supervisory training: An experimental study. Human Resource Development Quarterly 3(4), 345-356.

Shipper, F., & Neck, C. P. (1990). Subordinates' observations: Feedback for management development. Human Resource Development Quarterly, 1(4), 371-385.

Smith, R. M., Montello, P. A., & White, P. E. (1992). Investigation of interpersonal management training for educational administrators. Journal of Educational Research, 85(4), 242-245.

Sniderman, R. L. (1992). The use of SYMLOG in the evaluation of the effectiveness of a management development program. Unpublished doctoral dissertation, California School for Professional Psychology, Los Angeles.

Sogunro, O. A. (1997). Impact of training on leadership development: Lessons from a leadership training program. Evaluation Review, 21(6), 713-737.

Steele, R. W. (1984). The effect of interactive skills training on middle managers in a major corporation. Unpublished doctoral dissertation, George Washington University.

Tenorio, E. M. (1996). A level three evaluation of a management training program. Unpublished master's thesis, San Jose State University.

Tesoro, F. M. (1991). The use of the measurement of continuous improvement model for training program evaluation. Unpublished doctoral dissertation, Purdue University.

Tharenou, P., & Lyndon, J. T. (1990). The effect of a supervisory development program on leadership style. Journal of Business and Psychology, 4(3), 365-373.

Thoms, P., & Greenberger, D. B. (1998). A test of vision training and potential antecedents to leaders' visioning ability. Human Resource Development Quarterly, 9(1), 3-19.

Thoms, P., & Klein, H. J. (1994). Participation and evaluative outcomes in management training. Human Resource Development Quarterly, 5(1), 27-39.

Tracey, J. B., Tannenbaum, S. I., & Kavanaugh, M. J. (1995). Applying trained skills on the job: The importance of the work environment. Journal of Applied Psychology, 80(2), 239-252.

Tziner, A., Haccoun, R. R., & Kadish, A. (1991). Personal and situational characteristics influencing the effectiveness of transfer of training improvement strategies. Journal of Occupational Psychology, 64, 167-177.

Urban, T. F., Ferris, G. R., Crowe, D. F., & Miller, R. L. (1985, March). Management training: Justify costs or say goodbye. Training and Development Journal, 68-71.

Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. Personnel Psychology, 48, 347-375.

Werle, T. (1985). The development, implementation and evaluation of a training program for middle-managers. Unpublished doctoral dissertation, The Union Graduate School.

Williams, A. J. (1993). The effects of police officer supervisory training on self-esteem and leadership style adaptability considering the impact of life experiences. Unpublished doctoral dissertation, Georgia State University.

Wolf, M. S. (1996). Changes in leadership styles as a function of a four-day leadership training institute for nurse managers: A perspective on continuing education program evaluation. Journal of Continuing Education in Nursing, 27(6), 245-252.

Woods, E. F. (1987). The effects of the Center for Educational Administrator Development management training program on site administrators' leadership behavior. Unpublished doctoral dissertation, University of Southern California.

Yang, Z. Z. (1988). An exploratory study of the impact of a western management training program on Chinese managers. Unpublished doctoral dissertation, Pepperdine University.

Yaworksy, B. (1994). A study of management training in a criminal justice setting. Unpublished doctoral dissertation, Rutgers University.

Young, D. P., & Dixon, N. M. (1995). Extending leadership development beyond the classroom: Looking at process and outcomes. Proceedings of the Academy of Human Resource Development, 5-1.

**APPENDIX C**  
**CODING FORM**

**I. Study Identification**

1. Study ID # \_\_\_\_\_
2. Author(s): \_\_\_\_\_
3. Publication year: \_\_\_\_\_
4. Publication type:  
\_\_\_\_ 1) Journal article, book, or report      \_\_\_\_ 2) Dissertation or thesis  
\_\_\_\_ 3) Unpublished study

**II. Participant Characteristics**

1. Job classification level:  
\_\_\_\_ 1) Foreman/supervisor      \_\_\_\_ 2) Mid manager level  
\_\_\_\_ 3) Top management      \_\_\_\_ 4) Mixed levels  
\_\_\_\_ 5) Other (Please list: \_\_\_\_\_)
2. Organization type:  
\_\_\_\_ 1) Education      \_\_\_\_ 2) Government      \_\_\_\_ 3) Automotive      \_\_\_\_ 4) Military  
\_\_\_\_ 5) Medical      \_\_\_\_ 6) Financial      \_\_\_\_ 7) Technology      \_\_\_\_ 8) Utilities  
\_\_\_\_ 9) Manufacturing      \_\_\_\_ 10) Mixed      \_\_\_\_ 11) Unknown  
\_\_\_\_ 12) Other (Please list \_\_\_\_\_)
3. Country where program occurred:  
\_\_\_\_ 1) U.S.      \_\_\_\_ 2) Non-U.S.  
\_\_\_\_ 3) Both      \_\_\_\_ 4) Unknown

**III. Intervention Characteristics**

1. Program name: \_\_\_\_\_
2. Sample size: Experimental Group(s) \_\_\_\_\_ Control Group(s) \_\_\_\_\_ Total \_\_\_\_\_
3. Intervention type (check one):  
\_\_\_\_ 1) Formal training      \_\_\_\_ 2) Developmental relationships  
\_\_\_\_ 3) On-the-job experiences      \_\_\_\_ 4) Feedback programs  
\_\_\_\_ 5) Mixed
4. Content focus (check one):  
\_\_\_\_ 1) Problem solving and decision making      \_\_\_\_ 2) Strategic stewardship  
\_\_\_\_ 3) Employee performance      \_\_\_\_ 4) Human relations

\_\_\_ 5) Job and work redesign

\_\_\_ 7) Mixed

**IV. Outcome Characteristics and Effect Sizes (Check appropriate outcome level)**

- A. \_\_\_ **Learning-Knowledge and Subjective Learning**  
B. \_\_\_ **Learning-Knowledge and Objective Learning**  
C. \_\_\_ **Learning-Expertise (Behavior) and Subjective Behavior**  
D. \_\_\_ **Learning-Expertise (Behavior) and Objective Behavior**  
E. \_\_\_ **Performance - Financial (Results) and Subjective Results**  
F. \_\_\_ **Performance - Financial (Results) and Objective Results**  
G. \_\_\_ **Performance - System (Results) and Subjective Results**  
H. \_\_\_ **Performance - System (Results) and Objective Results**

1. Outcomes variables measured: \_\_\_\_\_

2. Instrument: \_\_\_\_\_

3. Instrument reliability: \_\_\_\_\_

4. Measurement method:

- \_\_\_ 1) Self perception      \_\_\_ 2) Self rating      \_\_\_ 3) Objective test  
\_\_\_ 3) Other person rating      \_\_\_ 4) 360 feedback  
\_\_\_ 5) Objective, not self-reported report  
\_\_\_ 6) (Please list \_\_\_\_\_)

5. Study statistics:

\_\_\_ A. Between group:

Experimental Group      Control Group  
N = \_\_\_\_\_      N = \_\_\_\_\_

Mean	SD	Mean	Mean SD	SD Difference	F or Diff	t-statistic	df	p-value
------	----	------	------------	------------------	--------------	-------------	----	---------

\_\_\_ B. Within group (Pretest-Posttest) or Pretest-Posttest With Control Group

Experimental      Control  
N = \_\_\_\_\_      N = \_\_\_\_\_

Mean	SD	Mean	Mean SD	SD Difference	F or Diff	t-statistic	df	p-value
------	----	------	------------	------------------	--------------	-------------	----	---------

Pretest \_\_\_\_\_

Posttest \_\_\_\_\_

\_\_\_\_ C. Other study statistics

t score	_____
F ratio	_____
p-value	_____
r correlation	_____
r-squared	_____
Chi Square	_____
z score	_____
p value	_____
df	_____

\_\_\_\_ D. Anticipated outcome: \_\_\_\_ Yes \_\_\_\_ No

## **APPENDIX D**

### **CODING INSTRUCTIONS**

#### **I. General instructions:**

1. Make selection based upon interpretation of information found in the study. If no category fits, check “other” and write a description of the variable on the line provided.
2. Check “unknown” when you are unable to determine the variable from information in the study. However, common sense and professional expertise should be used to make coding selections when the required information is not explicitly addressed by the authors but can be reasonably inferred from the report. Make a note concerning such decisions.
3. Check “mixed” when more than one of the selections is found in the study.
4. If you want to elaborate on coded information, asterisk the item and put a note at the bottom of the code sheet page describing the problem.
5. Please note that there may be more than one outcome in the study from which an effect size can be obtained. Use a separate coding sheet for each outcome and record one set of statistics from which to determine an effect size per outcome.

#### **II. Coding of Study Characteristics**

##### **Study Identification:**

1. Assign study identification number in order in which study is entered into the database.
2. Enter last name of all authors of the study.
3. Enter the year of publication.
4. Choose the type of publication from which the study was obtained. Select (1) if journal article, book, or published report; (2) if dissertation or thesis; or (3) if an unpublished study.

##### **Participant Characteristics:**

1. Choose managerial job level of majority of participants in study. Select (1) if focal group is supervisor or foreman level (entry level); (2) if mid manager level; (3) if top management level; (4) mixed levels; (5) other. Provide a description of “other” if chosen.
2. Indicate the primary organization type in which the majority of participants work.
3. Select (1) if program based in the United States; (2) if non US; (3) if both or multinational; or (4) if unknown.

#### Intervention Characteristics:

1. Indicate the name of the managerial leadership development program.
2. Enter sample size for study.
3. Select intervention type based upon the following definitions:
  - Formal training programs. Structured training programs in a formal setting either in the work place or in a location outside the organization that are designed to develop the individual employee are examples of formal training interventions.
  - Developmental relationships. Experiences in work settings where another individual influences the manager's personal development and relationships that individuals point to as their key sources of assessment, challenge, and support are considered developmental relationship interventions. Examples of these relationships include one-on-one mentoring, coaching, dyadic relationships or leader-match.
  - On-the-job experiences. On-the-job experiences stretch people, push them out of their comfort zone, and require them to think and act differently. This intervention includes an entire job such as redesigning a system or a piece of job, or serving on a temporary task force. Job assignments not only include work that is "assigned", but can also include an experience for which the manager volunteers. This category also includes development experiences with managers learning, growing, undergoing personal change as a result of the roles, responsibilities, and tasks they encounter in their jobs.
  - Feedback programs. Interventions are in the form of assessment data fed back to participants in a safe, supportive environment and a systematic collecting of opinions about a manager's performance from a wide range of co-workers, including peers, direct subordinates, the boss, the boss' peers, and people outside the organization (360 degree feedback). This managerial leadership development intervention includes assessment in formal feedback settings.
  - Mixed. Any combination of interventions as described above should be designated as mixed interventions.
4. Select content focus of the intervention based upon the following definitions:
  - Problem-Solving and Decision-Making - The emphasis focuses on generalized problem-solving and decision-making skills that would be applicable to a wide range of work problems that managers encounter.
  - Strategic Stewardship - Participants learn how to develop, manage, and lead an overall organizational program or sub-unit of an organization, including knowledge of trends or perspectives of the profession, the importance of innovation, how to transform organizational culture and values, implement and institutionalize change, communicate strategy and performance information to relevant subsystems, and understand organizational politics. The measurement of organizational outputs against strategic goals to promote continuous organizational improvement, the ability to acquire and allocate adequate resources conducive to achieving high performance is also included



in strategic stewardship course content. Participants may also learn to develop organizational systems (personnel, technology and control) that enable effective performance.

- Employee Performance – Participants discuss continuous improvement through feedback or the assessment of individuals’ knowledge, skills, and abilities. Participants can learn how to write performance objectives and performance appraisals that effectively contribute to the development of employees. Nurturing relationships through coaching and mentoring that bring out the best in employees and help them reach their highest level of achievement are examples of employee performance programs. These programs may also enhance participants’ chance of career success or enable them to adapt more quickly to the organizational culture. The experience may also teach participants how to develop clear employee goals, use motivation to build employee commitment, increase managers’ motivation, or modify a manager’s values or attitudes.
- Human Relations - Participants learn interpersonal skills that enable them to supervise, resolve personnel conflicts, collaborate, and communicate effectively. The focus is on human relations problems of leadership, supervision, attitudes toward employees, and communication. They learn to link the right employee to the right job, link human resource policies and procedures to job activities to enhance organizational and individual performance. Human relations programs help leaders and managers develop individuals into future leaders whose values closely match the organization, and to evaluate the organization’s policies and procedures in relation to employee performance. The ability to provide an open atmosphere is included in the course content.
- Job and Work Redesign – Participants learn how to redesign jobs that lead to employee satisfaction, create meaningful jobs, or develop job characteristics that motivate employees to do their best work. They learn to determine the appropriateness of forming work groups, to be proficient in team development, and how to enhance group dynamics to improve effectiveness in working together through confrontation and consensual resolution of conflicts. The quality of group decisions and verbal behaviors are covered in the job and work redesign content area.
- Mixed (General Management)– Participants learn topics in more than one of the categories above.

## V. Outcome Characteristics and Effect Sizes

(In this section, coder will determine outcome level for the analysis and will complete the remainder of the coding sheet based upon intervention outcome chosen. Please use the following definitions to choose the appropriate outcome category based on the Results Assessment System)

- Learning – Knowledge: the mental achievement acquired through study and experience.
- Learning – Expertise (Behavior): human behaviors having effective results and optimal efficiency acquired through study and experience within a specialized domain.
- Performance – Financial: the conversion of output units of goods/services attributable to the intervention into money and financial interpretation.
- Performance – System: units of mission-related outputs in form of goods/services, having value to the customer, that are related to the core organizational, work, process, and group or individual contributor to the organization.

Please use the following definitions to choose the appropriate outcome category based upon Burke and Day's model (1986):

- Learning – Subjective: principles, facts, attitudes, and skills learned during or by the end of training as communicated in statements of opinion, belief, or judgment completed by the participant or trainer.
- Learning – Objective: principles, facts, attitudes, and skills learned during or by the end of training by objective means, such as number of errors made or number of solutions reached, or by standardized test.
- Behavior (Expertise) – Subjective: measures that evaluate changes in on-the-job behavior perceived by participants, peers, or supervisor.
- Behavior (Expertise) – Objective: measures that evaluate changes in on-the-job behavior through a systematic process, with a supervisor or peer using a standardized instrument to complete the measurement.
- Results (Performance) – Subjective: organization results perceived by respondents, not reported by hard record, (e.g., subordinates' job satisfaction or commitment to the organization,) and group effectiveness perceived by subordinates.
- Results – (Performance) – Objective: tangible results, such as reduced costs, improved quality or quantity, promotions, and reduced number of errors in making performance ratings.

1. Describe outcome variables measured.
2. Enter name of instrument used to measure outcome identified.
3. Enter alpha coefficient of reliability from study.
4. Select method of measurement:
  - a. Self perception - perceptions of participants
  - b. Self rating – participants rate outcomes using a measurement instrument
  - c. Objective test - outcome measured through test scores
  - d. Other person rating – Evaluation completed by person other than the participant
  - e. 360 feedback – rating completed by follower and superior
  - f. Objective, not self reported – outcome determined through objective results, not reported by the participant

5. Enter quantitative statistics from study.
  - a. If the study used a posttest only with control group research design, complete section A.
  - b. If the study used a pretest posttest with control or single group pretest posttest research design, complete section B.
  - c. If correlational, or another design that provides additional study statistics was used, complete section C.
  - d. If the anticipated outcome was in the expected direction, check “yes”; otherwise, check “no.” A positive sign should be assigned to the effect size when the intervention group did “better” than the control group or the posttest score was “better” than the pretest score. Likewise, a negative sign should be assigned when the intervention group did “worse” than the control group. Of special note are situations where high scores on the dependent variable intentionally indicated worse performance, and the signs should be reversed in determining those effect sizes.

## VITA

Doris Bowers Collins was born in Petersburg, West Virginia and received her early education in the public schools in Pendleton County, West Virginia. She graduated from Franklin High School in 1962. She was awarded a Bachelor of Science degree in secondary education in 1966 and a Master of Science degree in biology in 1970 from West Virginia University.

She was employed as Assistant to the Dean of Women and Director of Women's Housing at Georgia Institute of Technology from 1970-1977. She served as Assistant Director of Housing at Louisiana State University from 1977-1982; Associate Director of Housing from 1982-1993; and, Director of Residential Life from 1993-1999. During the dissertation process, she was employed as Associate Vice Chancellor for Student Life and Academic Services at Louisiana State University, with responsibilities for the LSU Union, Recreational Sports, Residential Life, Student Media, and Child Care.

She is a nationally recognized consultant in college student affairs, has served as president of the Association of College and University Housing Officers-International (ACUHO-I), participated in national teleconferences, and chaired national committees and task forces on issues related to services and programs offered on college campuses. She is the recipient of numerous national and regional leadership awards. Her publications on program evaluation, organizational performance, and high performance leadership competencies can be found in *Human Resource Development Review*, *Journal of Leadership Studies*, *Cutting Edge Papers of the Academy of Human Resource Development* and *Advances in Developing Human Resources*.